Applications and Industry

 November 1957

Transactions Papers

NUMBER 33

Published Bimonthly by

AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS

(See inside back cover)

Applications and Industry. Published bimonthly by the American Institute of Electrical Engineers, from 20th and Northampton Streets, Easton, Pa. AIEE Headquarters: 33 West 39th Street, New York 18, N. Y. Address changes must be received at AIEE Headquarters by the first of the month to be effective with the succeeding issue. Copies undelivered because of incorrect address cannot be replaced without charge. Editorial and Advertising offices: 33 West 39th Street, New York 18, N. Y. Nonmember subscription $5.00 per year (plus 50 cents extra for foreign postage payable in advance in New York exchange). Member subscriptions: one annual subscription in consideration of payment of dues without additional charge to any one of three divisional publications: Communication and Electronics, Applications and Industry, or Power Apparatus and Systems; additional annual subscriptions $2.50 each. Single copies when available $1.00 each. Entered as second-class matter at the Post Office at Easton, Pa.

Printed in United States of America

Number of copies of this issue 12,300

Ground-Power Equipment for the Support of Air Force Weapons

T. B. HOLLIDAY
FELLOW AIEE

THE MATERIAL that is presented in this paper was gathered during an engineering study on the same subject which was made for the Air Research and Development Command of the United States Air Force. The objective of the study was preparation of recommendations for needed development in the area of ground support for secondary power. In aviation terminology, secondary power is defined as that power which is used for purposes other than the propulsion of aircraft and missiles. Aircraft and missiles are defined as weapons, and a weapon system includes everything that is needed for maintenance, calibration, testing, and training during operational use of the weapon. "Ground power" substitutes for air-borne secondary power during tests. The familiar secondary power systems are electric, hydraulic, and pneumatic. During the study, it was found desirable to treat air conditioning as a fourth and newly important type of secondary power.

Comments were obtained from all combat and technical Commands of the Air Force, from all weapon manufacturers who serve the Air Force, from engine manufacturers, from manufacturers of secondary power-generating equipment, and from some manufacturers of equipment that uses secondary power. In all, more than a half million items of information were gathered. The summary and collation of these data revealed a pattern of future needs and suggested desirable solutions to those needs.

In order to make a prediction of the probable needs in ground power, it is first necessary to predict the trends of the air-borne secondary power systems. These trends cannot be estimated without first making some assumptions, if not predictions, as to the trend of aircraft and missiles. This was the approach that was used during the Air Force study, and it produced much information of indirect value, which was to contribute to the planning of ground power. The results as they pertain to secondary power and the support thereof are presented here.

This article is limited to a discussion of ground electric power. It was found that weapons of the future can be served by five families of power units: (1) ramp unit, (2) trailer-mounted mobile unit, (3) stationary or portable unit, (4) fixed base power unit, and (5) electric motors. Of these, all but one are primarily sources of electric power. The second item will be found most useful as a hydraulic power unit. It became apparent that the turbine prime mover will rapidly supplement and then supersede the reciprocating engine. Thus, it was concluded that future developments should be based on the turbine drive.

There are two primary problems in any effort to plan ground power support. First, the effectiveness of the weapon must not be impaired. Second, and to save costs, as many common items as possible should be developed to conserve development time, personnel, and funds.

General Predictions

About the only safe prediction that can be made with regard to aviation development is that any given period will yield about twice the technological development which was accomplished during the same length of time in the immediate past. It is becoming increasingly apparent that there is no limit to the flight altitude and speed that will be achieved. The thermal barrier will be overcome by developments in materials, or by flight at higher altitudes. The developments in missiles will lead the way into outer space, but soon afterward, the missile will be converted to a manned aircraft of essentially similar performance.

Flight into outer space will advance the limits of environmental conditions. Higher flight speeds brings increased aerodynamic forces and the need for greater power in secondary power systems for the servo controls. Contributing to the achievement of higher speeds will be more compact designs of fuselage and wings. Yet more and more secondary powered equipment will be installed in the reduced space. Heat losses from that power must be carried away. The result indicates a much greater importance for air conditioning in the future. Development of components that can operate in higher ambient temperatures will reduce the power required for air conditioning, but the need will probably continue.

The higher speeds of the future will bring new problems in precision. Navigation, whether by human skill or by missile robot, will need greater precision. The necessary accuracy starts in the secondary power system, particularly in the electric system. It is evident that there will be a constant demand for more and more preciseness of control in the future. It is not likely that the aviation industry will ever be satisfied, because still more accuracy will always be sought in guidance equipment. Additional and supplementary precision is now built into the electronic subsystems. If more accuracy can be provided for the input secondary power system, less will have to be added by electronic means. Heretofore, only the electric system has been the subject of the demand for greater precision, but it is likely that other types of secondary power will tend toward greater precision in the future.

The hydraulic system will be the next to feel the pressure toward better control. Hydraulic power is the ideal means for servo control, as it provides ample power with light weight, is well suited to exerting lineal forces, is reasonably fast in response, and is not resilient or soft. Electric devices do not lend themselves to lineal actuation through any but very short distances, and most of them are too slow to be used as servo controls. The pneumatic system has too much resiliency to be used in continuous-duty servo applications, although it is very good for one-time applications.

The transistor will reach a major plateau of its usefulness by 1962. It has not brought all the blessings hoped for.

It will permit a great reduction in the size and power consumption of many electronic systems; particularly digital computers. This will reduce the amount of air conditioning that is needed to remove electric heat, but it will demand more precise temperature control. Many transistors have given characteristics over a small range of temperature. To achieve the ultimate precision in the electronic system, it will be necessary to control temperature around the transistors to quite narrow limits.

The power rating of the electric system

Paper 57-500, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE East Central and Middle Eastern District Meeting and Air Transportation Conference, Dayton, Ohio, March 7–9, 1957. Manuscript submitted January 28, 1957; made available for printing April 8, 1957.

T. B. HOLLIDAY is with the Holliday Company, Elmhurst, Ill.

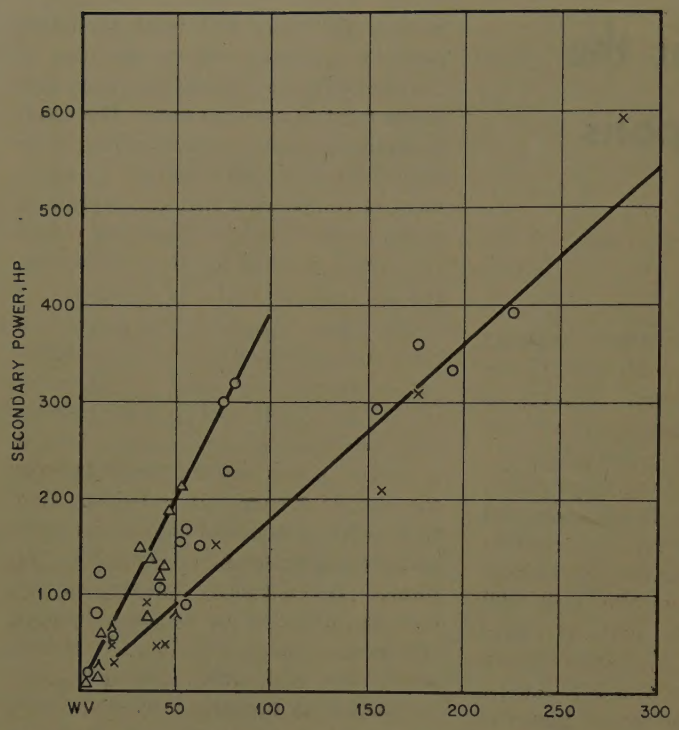


Fig. 1. Total secondary power load with respect to product of weapon gross weight and speed in Mach number

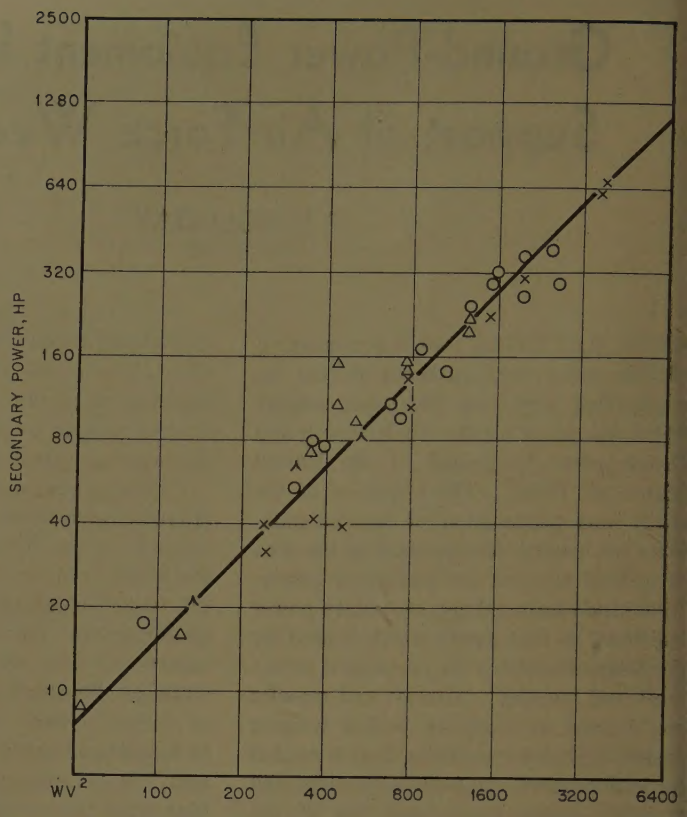


Fig. 2. Total secondary power load with respect to product of weapon gross weight and square of speed using indicated air speed

will continue to grow despite the advent of the transistor, but its advance in size will not be as dramatic as in the last 10 years, nor will it keep pace with the growth in weight and speed of aircraft. Relatively, the electric system will lose ground in quantity of power.

The hydraulic system will have a rapid growth in the next 10 years. Aircraft are now reaching speeds at which manual control is out of the question. Power boost must be used. Until now, it has been possible to use aerodynamic forces to counterbalance control forces, so that manual effort could control the flight. Power boost has been added to make the job easier. Now that aerodynamic forces are too large to use in this manner, and since power boost is always available, the aircraft designers may elect to design instability into the aircraft deliberately in order to attain still higher performance. This will increase the power rating of the secondary power system, and will force installation of a dual system in many types of aircraft, in order to have a safety factor for emergencies. A tenfold growth of hydraulic installed power in the next 10 years is predicted. Much of this will be due to the increased weight and speed of the aircraft, and some from increased uses for hydraulic power. Where three or four hydraulic subsystems have been used in the past, e.g., wheel brakes, landing gear retraction, flap actuation, and autopilot servos, as many as 12 can be expected in the next 10 years. Some of these will be continuous-duty applications in which hydraulic motors do work similar to that done by electric motors.

Low-pressure pneumatic power has been tried in the past 5 years and seems to be settling into a place of usefulness. It is tempting to use bleed air from the turbine engines through a pneumatic coupling to drive generators of secondary power. Although it is not an efficient coupling, it does remove accessories from the engine and permits a reduction in the frontal area of the engine. Advantage could also be taken of the ducts for deicing air as a path for pneumatic power. But, it is likely that deicing will be omitted on any aircraft that can cruise at altitudes beyond icing conditions. Todays's piston engine aircraft which are not able to evade all icing conditions, use deicing equipment for less than 3% of the total flight hours. Military aircraft need only pass through icing conditions enroute to or from iceless environment, and do not need deicing equipment at all. This means that the deicing ducting which could convey pneumatic power will not be installed. It is anticipated that the low-pressure air-turbine-driven devices will be restricted to emergency power generators and to missiles. The free turbine drive within missiles can use hot gases bled from the rocket engine, or ram air pressure.

High-pressure pneumatic power can be used in many special one-shot jobs, and as stored energy in the smaller missiles. This is a very economical way in which to carry stored power. The maximum pressure in such storage is likely to exceed 10,000 psi (pounds per square inch) in the next 10 years. The foregoing has given some general predictions as to air-borne secondary power systems. Some detailed characteristics are now considered.

The air-borne auxiliary power unit will have new importance in the future. Trends in weapon propulsion show that a static device may be the source of thrust in future designs, thereby removing the source of rotary power for electric generators. The rocket engine in missiles and the experimental aircraft is a forerunner of this trend. The increased use of the ram-jet engine is another. Space travel will bring entirely new operational concepts. A separate and independent source of electric power is and will be needed in such aerial vehicles.

Detailed Predictions

It was found that the total of secondary power has a relationship to the weight and speed of the weapon. In military aircraft, there is a small amount of power, as for radio communications, that may be said to be common to all aircraft without

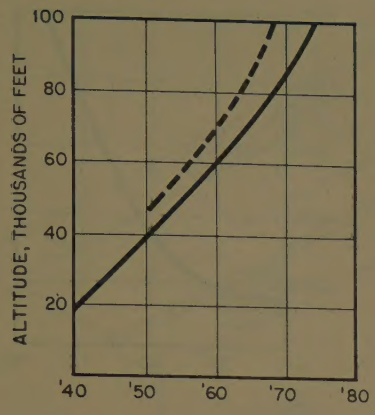


Fig. 3. Trend of operating altitude for manned aircraft (solid line) and air-breathing missiles

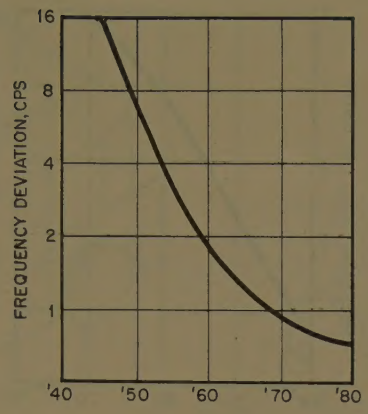


Fig. 5. Trend in reduction of frequency deviation (half of total tolerance)

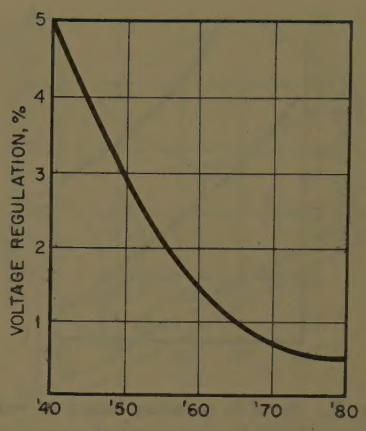


Fig. 6. Trend in reduction of voltage regulation (half of total tolerance)

regard to size. However, the amount of such power is almost negligible and is considered to be a part of the total for this estimate. The total power that is to be considered includes electric, hydraulic, pneumatic, and air conditioning. Any attempt to predict the amount of any one type of secondary power will result in confusion, because there are many accessory tasks that can use different types of power, and the various weapon manufacturers will choose different types of power for these. For example, the motor that drives an air-borne electric inverter can be electric or hydraulic.

It is necessary to consider only the power delivered by the secondary power generators. An evaluation that is based on the input power to these generators would be made erratic by the differing efficiencies of driving couplings. For example, two alternators might deliver 75 kw, which is an output power of approximately 100 hp (horsepower). A direct driven alternator might deliver this output with an input of 125 hp. The addition of a variable speed drive could demand an input power of approximately 140 hp, and the use of a pneumatic coupling to a driving turbine could require still higher input power. To avoid these several choices, only the output power of the secondary generators and pumps is considered.

The first attempt to make this analysis used secondary power as a function of the product of weapon weight and speed. The equation

$$P = kWV$$

where k is a constant, was used. The results of a plot of many aircraft is shown by Fig. 1. The ordinate of this plot is the total secondary power in horsepower. The abscissa is the product of gross weight in thousands of pounds and of velocity in Mach numbers. Two paths of correlation appear. Most of the supersonic aircraft lie on the upper line. It is obvious that this approach is not satisfactory, and that secondary power is dependent on an exponent of the velocity factor.

To bring better correlation, three steps were taken. First, indicated air speed was used instead of Mach number; second, the square of the velocity was used; and finally, some allowance was made for differing practices among weapon manufacturers with regard to hydraulic servo control systems. When dual systems were installed, the power for only one of them was counted. These changes brought the two curves together as shown by Fig. 2. This correlation is a reasonably straight line that can be projected into the future with some degree of certainty. This prediction of total power is important in the planning of ground power units, because the power rating of the servicing power unit will have some relation to the installed secondary power. After studying the highest single loads of the systems combined with the air-conditioning requirements that would accompany testing of those loads, it was decided that the ground power electric servicing unit would need to have about half the power rating shown for the total secondary system.

The air-conditioning loads of commercial transports cause their position on the plot to be high in the case of the older designs. As we move into the era of jet transports, the secondary power requirements come into line with those of military aircraft. This shows that the human factor becomes of less importance with higher performance in so far as air conditioning is concerned. In the future, the air-conditioning needs of the weapon itself will far overshadow those for the personnel in the weapon.

The operating altitudes of weapons directly concern the environment for which secondary air-borne power equipment must be designed. These are of indirect interest in the ground power unit, for it will often be required to drive air-borne generator units in the servicing of the weapon. A projection of the trend in operating altitude of weapons is shown by Fig. 3, where the solid line represents aircraft, while the dotted line represents missiles. This plot shows an operational

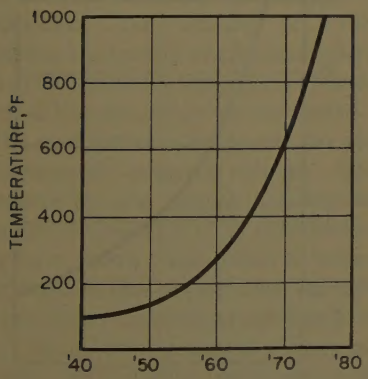


Fig. 4. Trend of environmental temperature around some items of air-borne electric equipment

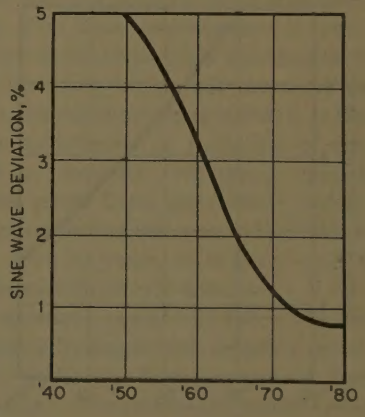


Fig. 7. Trend in reduction of factors that affect wave shape

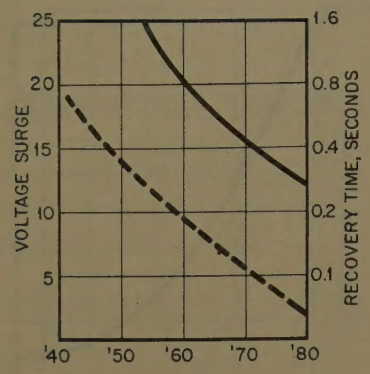


Fig. 8. Trend in reducing factors of voltage surges. Solid line is amount of surge in either direction and dotted line is time to recover to regulated values

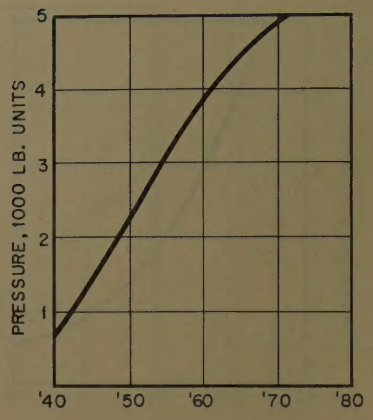


Fig. 10. Trend in pressure of air-borne hydraulic system

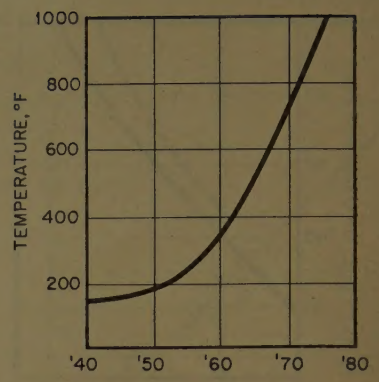


Fig. 11. Trend in hydraulic fluid temperature for normal operation

prediction, and not that of test and experimental flights and weapons. The experimental flight of today may be operational in another 10 years. It is already being required that air-borne secondary power generators be designed to be free of environment, and it is apparent that many more items of the secondary systems must follow suit in the next 20 years.

The environmental temperatures around critical pieces of equipment, particularly generators and pumps, will increase steadily, approximately as shown by the chart of Fig. 4. The temperature environment for most other items of secondary equipment will be less severe. Many of them will be located in space that is air conditioned. Others will be installed in space that is not close to heat generating sources of power.

Electrical Characteristics

The problems facing the designer of air-borne electric-power-generating equipment can be summed in one word, "precision." For basic power, the accepted standard is 3-phase alternating current at 208/120 volts and 400 cps (cycles per second). Generation of d-c power is rapidly fading out, although the need for direct-current is not. The d-c power that is needed will be delivered by transformer-rectifier units. There are some interesting possibilities for the direct generation of electric power, and they are likely to produce direct instead of alternating current. A technical breakthrough of this nature could radically alter air-borne electric systems.

Some consideration is being given to higher voltage systems that may be needed in larger aircraft of the future. The industry has tentatively agreed on a 3-phase 440-volt system for the proposed applications. There are several applications for other and higher frequencies, and they are brought about by the need for faster response from magnetic amplifiers. These frequencies have included 800, 1,200, 1,600, 2,000, 2,400, and 4,800 cps. If such special frequencies continue to be needed, it should not be a multiple of the basic power frequency, 400 cps. One frequency that should be avoided is 2,400 cps, for that is the ripple frequency of a 3-phase transformer rectifier unit. Successful development of power transistors that can replace the function of magnetic amplifiers will reduce or eliminate the need for special frequencies.

The preciseness of frequency control was one of the great debates encountered during the survey. There are four schools of thought. The first believes there is limited application on turboprop engines for variable-frequency alternators to supply raw or basic power. The second calls for a system basic power in which the frequency is controlled in a total range of 20 cps, 400 ±10 cps. The third group wants better control, and it is the prediction of this group that is shown by the plot of Fig. 5. The graph shows the permitted variation in each direction, half the total variation. All of these groups are fully aware of the need for very precise frequency control in critical electronic systems, but they believe that the refinement can and should be done in the power supply of the electronic equipment. The group contends that it is just as easy to refine all of the power at the outset, rather than a part of it.

Where the drive for precision in frequency will end is not easy to predict. Temperature-controlled crystals can deliver a precision in the order of 10^{-8}, which means of precision of 0.25×10^{-6} cps. The demand of the precision school today is for 0.01 cps, but there is little reason to doubt that the ultimate in precise control will be wanted eventually. The extreme precision is needed in missile guidance systems which cannot be countered. Such precision in a reference signal from the electric system can take the place of other references and is much easier to adapt to the guidance equipment.

Greater accuracy in the control of voltage will also be sought. The goal of weapon manufacturers is shown by the Fig. 6 graph. Such precision can be realized at only one accurate voltage control is not high, and is not a factor.

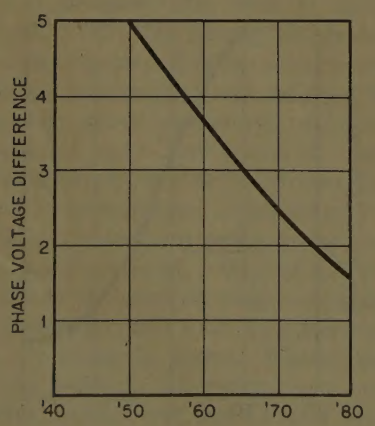


Fig. 9. Trend in reduction of differences between phase voltage of 3-phase system

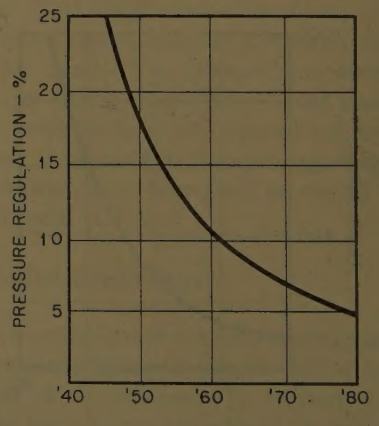


Fig. 12. Trend in pressure regulation (total tolerance)

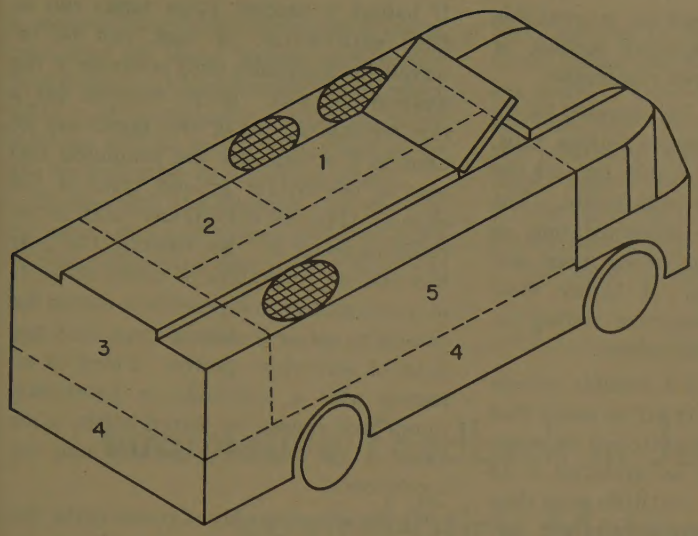


Fig. 13. Suggested design of ramp vehicle to provide electrical power to weapons

1—Turbine prime mover
2—Gear box and driven units such as alternator, generator, and pumps
3—Control space
4—Tanks for various fluids and ballast
5—Air-conditioning unit

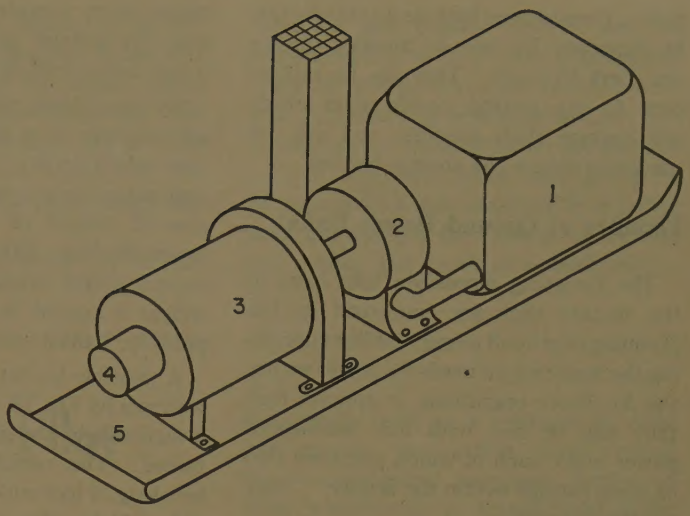


Fig. 14. Suggested arrangement of stationary electric generating set

1—Turbine drive
2—Free piston compressor
3—Alternator and gear reduction
4—Exciter
5—Mounting base

Wave shape of the output voltage will be more accurately limited in the future. The trend in the reduction of any deviation from a sine wave is shown by Fig. 7. These reductions apply to harmonics, single or multiple, and to crest factor. They are of particular importance to the ground power unit since it must match the quality of the air-borne units. Modulation is an undetermined factor in this trend at this writing. Modulation can occur to both voltage and frequency. The variable speed drive is a major source of modulation, particularly in voltage and to a lesser extent in frequency. Oscillations within the drive can vary frequency at about 1 cps per second, and apparently this is not detectable by the instruments that are normally used. Better instrumentation will permit its detection and may also lead to a campaign to reduce the modulation. The matter of exactly matching the quality of air-borne alternators could force the adoption of variable speed drives on ground power units.

The modulation of frequency can cause a considerable pulse of voltage. Measurements by one weapon manufacturer with the co-operation of an electrical supplier have shown these pulses to be more than 10% of line voltage, far exceeding the permitted variation of regulated voltage. Again, these pulses of voltage have not been the designated cause of any difficulty in load equipment, but it is feared that if instrumentation is made available that will detect the pulses, there will follow an effort to reduce or remove the pulses

Recovery from surges caused by the addition or subtraction of unit load is also scheduled to tighten as shown by Fig. 8. A single surge of power in an airborne system is not particularly disturbing but periodic surges can be very disturbing. A single surge will throw the pip of a radar scope off limits, but it soon recovers. Repetitive surges can be troublesome. In missile flight, the electrical loading can be scheduled so that surges are not troublesome.

As more experience is gained with 3-phase electrical systems, more of the individual loads go from single to 3-phase operation. The weapon manufacturers are very careful to balance single-phase loads by distribution among the three phases. With this progress comes a demand for less and less difference in phase voltages. The trend is shown by Fig. 9.

Other Secondary Power

The comparable trends in hydraulic systems will be of interest to the electrical engineer, particularly one who is concerned with ground power. The trend in system pressure is shown by Fig. 10. The greater power demands that are coming in the next 10 years would justify a still higher growth in pressure, but the zone of diminishing returns is reached at about 5,000 psi and even less in smaller aircraft. Beyond this critical value, the weight of heavier tubing offsets potential gains.

The trend in fluid temperature within the hydraulic system is shown by Fig. 11. The growth in temperature is due to two factors; the demand for more power, and the proximity to sources of great heat. Silicone fluids will permit a rise to about 800 degrees Fahrenheit from the current values of 275 degrees Fahrenheit. To achieve the maximum limits that are forecast, it is likely that lessons will be drawn from nuclear experience to use liquid metals as the fluid. Hydraulic fluids face another hazard of the nuclear era. Neutron radiation will damage many of the hydraulic fluids that are known today.

There will be an increasing demand for more accurate control of system pressure, and the possible trend is shown by Fig. 12. This requirement is not as severe as that which has been imposed on the electric system. At the same time that the regulation is reduced, it is likely that suppression or reduction of ripple will be required. As pump operating speed grows from the current rating of approximately 4,000 toward 12,000 rpm in the next 10 years, it is likely that the amount of ripple will decrease.

Pneumatic power will follow the same trends as hydraulic with regard to pressure and ripple in high-pressure air-borne systems. The pressure in stored systems is likely to more than double those for operating systems in the same time pe-

riod. Pressures as high as 12,000 psi can be expected for use in missiles during the next 10 years. This can be important to the ground power units which will service such missiles, and will be used to recharge the storage bottles.

Families of Ground Power Units

The foregoing discussion lists some of the factors that are concerned in the planning of ground power. After appraising the many major needs for power within the Air Force operations, it appears that they can be met with five families of power units, each of which will need two or more ratings within the family. They are the following:

1. A self-propelled ramp vehicle to supply electric (and other) power for servicing and maintaining aircraft or missiles. This vehicle can take two forms; the first a conventional vehicle that can be transported in cargo aircraft, and the second, a "pod" mounted on but separable from the tactical aircraft that it supports.

2. A trailer-mounted unit for nonperiodic use. These will be single purpose units as for hydraulic power or air conditioning.

3. Portable (skid mounted, stationary) electric generating sets of higher power rating than that of the mobile vehicle.

4. Electric motor drives.

5. Fixed power plants, electric, of rather large ratings to service detection and missile control centers.

The first family has been of greatest interest to the Air Force and to weapons manufacturers. The concept of a ramp vehicle has grown steadily in recent years, and several designs have appeared. Today we are faced with the need for still more power from such a power unit, and we have a chance to do some planning around a turbine prime mover within that vehicle. The turbine is ideally suited to driving an alternator at constant speed, as it is essentially a constant-speed device itself. Since a turbine is used, it gives a capability of delivering low-pressure pneumatic power in considerable amounts, and this bleed air can be used for other purposes. Today this air can be used for pneumatic starters on the weapon engines, for ground testing of free-air or ram-air driven power units within the weapon, and as a source of pressurization and of ventilation, since the output air is harmless.

When electric power is delivered and used within a weapon, the excess heat, virtually all of that power, must be withdrawn by means of cooling air. This means that the power unit must have an air conditioner to generate that cooling air. The air conditioner can use air cycle techniques since compressed air is available from the turbine compressor section, or it can use a shaft-driven compressor.

Self-propulsion will be provided by an air coupling to a separate turbine unit. Four-wheel drive is desirable, because the unit can serve as a tow vehicle also. The over-all weight of a combined unit of approximately 250-hp rating need not exceed 4,000 pounds. If higher dead weight is needed to improve towing capability, ballast can be added.

A possible layout of a suitable vehicle is shown by Fig. 13. It will be noted that a surprisingly low silhouette can be maintained. The vehicle as sketched is 15 feet long, 6 feet wide, and little more than 4 feet high. The secondary power section of the vehicle has been located at the rear of the unit, to make it easier to meet the needs of the weapon system concept. This concept as practiced by the Air Force and the aviation industry has accelerated the development of advanced weapons. It is very jealously guarded by both the military and the industry, and yet both agree that a reasonable amount of standardization is to be desired. For that reason, those elements of the ground power unit that are concerned with matching the secondary power system of a weapon exactly are kept to the rear of the unit. The four wheels, propulsive means, air conditioner, and turbine drive are not concerned with any particular weapon, and are portions of the vehicle that can be standardized. The secondary power alternator (and generator, if used) are directly involved in the matching of air-borne characteristics, and they are located to the rear of the vehicle where they can be readily interchanged. Some designs of alternators do not use Army-Navy standard mounting flanges, and it is proposed that the reduction gear box which receives power from the turbine and drives the alternator, also be readily removable. Now, by changing a module at the rear of the vehicle to another similar unit that matches the airborne equipment, an exact matching of the air-borne system can be made. Associated equipment and controls should also be removable through the modular concept of design.

In the sketch of Fig. 13, the space 1 has been allocated for the turbine prime mover. Inlet and exhaust are on the upper surface of the vehicle. Space 2 is for the reduction gear box and driven devices, alternators, generators, pumps, and air-conditioning compressor, as needed. Space 3 is for the control of the secondary power system. Space 4 represents ballast tanks which can carry a variety of fluids. Fuel for the turbine is an obvious need. If ballast is needed, other tanks can be filled with water. A tank can be reserved for hydraulic fluid reservoir if the weapon needs it. If the weapon has a ram-jet engine, one or two tanks can be used as a storage for the simulated fuel that is needed for ground tests of the engine. The air-conditioner unit is located at space 5. By treating the gear box and its generators of power as a removable module, it is possible to design the vehicle to use or to deliver more than one type of secondary power. There is no reason why a hydraulic or pneumatic pump unit cannot be driven at the same time as the electric generator and air conditioner.

In the second family of power units, the first type that is likely to reach power values which justify a turbine drive is the hydraulic power unit. Within 10 years, a unit to deliver 50 gallons per minute at 5,000 psi will be needed. This will require a 200-hp prime mover. It should be designed with the module concept also, because it is apparent that in the rapid development of hydraulic systems, many pumps and fluids will emerge. One weapon may have systems of differing pressure ratings, and systems that use different fluids. After 10 years, there may be a need for external cooling during the ground checking of hydraulic systems, and this will require an air conditioner unit as well.

A suggested form for the stationary power unit of the third family is shown by Fig. 14. Again, the turbine is selected as the prime mover. This type of unit may be the successor to the familiar diesel-engine-driven alternator. As shown, the unit is driving a 400-cps alternator, and this is to be recommended for all Air Force power around weapons and their associated controls. In this type of unit, fuel economy is to be sought, since it will be called on to operate continuously. It is suggested that the unit be able to operate without shutdown for 9,000 hours, or one year, to meet Air Force needs in its remote stations.

A free piston compressor appears to be a promising source of the hot gases to drive this turbine. It can use a great variety of fuels, and with diesel oil, it is able to exceed the performance of the diesel engine in regard to fuel economy. A very conservative design of a 500-kw unit to deliver power at 400 cps should not weigh more than 3,000 pounds. The combination should offer fuel economy, long life between shutdowns and overhaul, and weight low enough to make transportation by air very easy. Among the disadvantages, this compressor is still

a reciprocating device, although it is simpler than its diesel engine counterpart. It also needs cooling, usually liquid, which adds to its complexity. Filtering of the inlet air will be just as important as for the reciprocating engine it supplants.

In line with the thought that the stationary power unit deliver 400-cps power, it is also suggested that the last family to be considered in this paper, the electric-motor drives, also be designed to use 400-cps power. They can use higher speeds and this, with the higher frequency, will give great reductions in unit weight. The higher speeds are more suitable to the driving of many of the pump units that are forthcoming.

The fixed power plants for missile detection and control centers, are too large a subject to be a part of this paper. They too will present some notable problems in quality of power, reliability, life, and security. Ground power for the support of weapon systems has grown in 20 years from a unit comparable to power lawn mower to a fixed power plant that can supply the needs of a small city.

Informative Feedback in Jet-Pilot Control-Stick Motion

N. D. DIAMANTIDES
ASSOCIATE MEMBER AIEE

Synopsis: In many closed-loop systems in which a human operator forms a link, the operator adds a special component to the error signal in continuously monitoring the system transfer function. Such a signal component, known in cybernetics as informative feedback, was found to be characteristically present in the control-stick motion of jet-airplane pilots. It was found, furthermore, that the amplitude of this feedback signal is subject to a learning process and that the over-all phenomenon, which falls within the scope of both control engineering and psychology, is amenable to quantitative analysis.

THE DESIGN of any block in a closed-loop servo system is dependent on all the blocks of the system. Wherever a human being is involved, as in a manned control system, the over-all design is generally prejudiced by lack of precise information on the human block. It is in this light that the increasing studies of the human organism in terms of a transfer function or, since nonlinearities are involved, a "describing function," which recent years have witnessed, are fully justified. The possible usefulness of such a concept increases yearly with the difficulty of the human operator's role in handling machines of greater complexity and speed. Rising costs and added safety hazards of such machines also heighten the need for more accurate performance prediction prior to construction.

In the specific case of piloted high-speed aircraft, the availability of an adequate analog of the pilot would provide a highly desirable tool for both airframe and flight-trainer design.

A number of pilot simulators have been built and tried both in the United States and abroad. However, the novelty of the art and limited knowledge of the human organism, viewed as a highly nonlinear and extremely complex mechanism, have hindered the success of simulating the physiological and psychological functions of the prototype. Nevertheless, in combination with accepted statistical methods they constitute probably the only effective means for evaluating the adequacy of the mathematical models constructed to represent analogs of human functions. This paper deals with one specific aspect of the human-analog study: the phenomenon of informative feedback in pilot control behavior.

An airplane simulator designed at Goodyear Aircraft Corporation for pilot control-efficiency evaluations made possible the observation of a highly interesting feedback phenomenon in pilot control-stick hand action. This action was evoked as follows. The pilot was placed in a mock-up cockpit whose horizontal position was disturbed by a simulated wind. The pilot's task was to keep the ship flying level at a constant altitude and speed. The action thus evoked was, for practical purposes, similar to what would occur in actual flight under similar atmospheric conditions. Methodologically the investigation was based on a parallel study of the performance of the human operator and his analog, which was built of electronic components on an analog computer.[1,2] To avoid the possibility of physiological crosstalk, only one degree of motion, pitch, was simulated.

Paper 57-494, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE East Central and Middle Eastern District Meeting and Air Transportation Conference, Dayton, Ohio, March 7–9, 1957. Manuscript submitted July 24, 1956; made available for printing February 21, 1957.

N. D. DIAMANTIDES is with the Goodyear Aircraft Corporation, Akron, Ohio.

The airplane mock-up and the original version of the pilot simulator were designed by R. J. Mead, who is now with McDonnell Aircraft Corporation. The mock-up was contractually supported by the U. S. Navy, Bureau of Aeronautics. An improved version of the simulator, a joint effort of the author and R. J. Mead, was used in this study. Dr. A. J. Cacioppo introduced the psychological approach to the problem. It should be noted, however, that the writer accepts sole responsibility for the conclusions reached in this paper.

Principles

A neuromuscular action such as the motion of a pilot's hand on the control column must be initiated by a command from the central nervous system. In the pilot and simulator situation such a command was evoked by the amplitude and rate of motion of the artificial horizon and by the acceleration of the seat. These visual and motional stimuli informed the pilot that a disturbance was taking place that he was to correct by making the proper control-stick motion.

Such a signal flow from the sensor units through the central nervous system to the motor units is influenced by the two well-known physiological discontinuities of threshold and reaction time. These two inherent characteristics of the human servo act as a gate in the path of the information transfer between input and output.

The input signal received by this servo is further modified by other characteristic parameters of the system. Such parameters are attributable to both the neurophysiology of the central nervous system, and the nature of the sensor and motor elements. Their effect is a proportional and derivative one (rate). For instance, when the apparent movement of the actual or artificial horizon indicates airplane displacement, the eyes send a signal to the brain that is related to the amplitude of the displacement. Acceleration will cause a signal to be generated in the vestibular organs of the

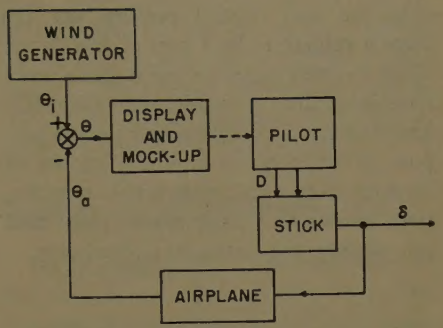


Fig. 1. Block diagram of closed-loop system

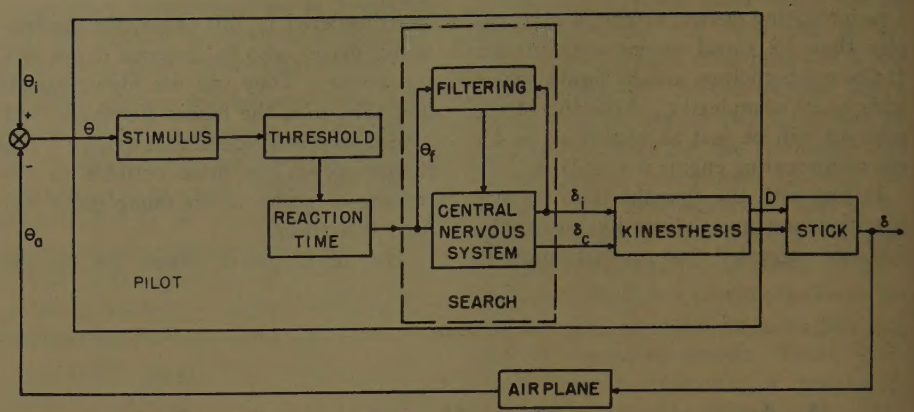


Fig. 2. Block diagram of physiological closed-loop system

inner ear and the proprioceptors of the muscles, tendons, and joints. Rate information is provided by the ear canals through endolymph flow, and by the eyes through peripheral vision resulting from the successive stimulation of the rods of the macula. In the case of the motor elements, it is known that the muscles involved contract or relax and at the same time move with their supporting bones in carrying out the commands of the brain. Consequently, it is seen that displacement and its time derivatives are important as causes of inertial forces, muscle elasticity, and fluid damping.

Apparatus

The apparatus used in the study of the feedback occurrence in question was simply a pilot seat mounted rigidly on a platform hydraulically driven about a horizontal axis. A visual display and a control column were placed on the platform. The display consisted of a 5-inch laboratory cathode-ray tube masked to the size and shape of a gyro-horizon instrument. The trace was deflected vertically as the platform was moved. In this manner the movement of the seat during actual airplane pitch as well as the related movement of the artificial horizon could be represented. Only pitch-control investigation was attempted to restrict the measurements to a single degree of freedom.

A 24-inch length of pipe having an actual airplane-handgrip served as the control column. The column could be moved about a horizontal axis located under the platform floor and was made to work against an inertial load, an adjustable spring, and viscous-friction forces. A damper cylinder with a control valve produced the viscous damping. Control-stick motion was translated into a proportional electric signal by a potentiometer, which was located under the platform. This signal was fed back into the mock-up and display to overcome the original disturbance.

Fig. 1 shows the closed-loop system as formed by its four basic components of mock-up and display, pilot, stick, and airplane. Since the three nonhuman components were defined in advance, the problem was to study the make-up of the pilot block. The role of input D is discussed in the appendix.

For the study, an electronic analog of this physical loop was constructed of operational and recording units of a GEDA* analog computer. In the analog the display's function of supplying the stimuli was performed by a simulator. Electronic analogs of the control column, the airplane, and the human operator, the latter perceiving and correcting for disturbances, were incorporated. The fact that the task of this study was simple enough as to be practically automatic made this last device feasible.

Fig. 2 translates Fig. 1 into physiological terms. The basic principles involved are discussed in the appendix. It should be noted that random noise effects are neglected and that the pilot block includes search in addition to the basic processes of stimulus, threshold, reaction time, and kinesthesis (neuromuscular action).

Search was introduced to enable the simulator to match the performance of jet-interceptor pilots. When a pilot was called upon to keep his plane flying level against random pitch disturbances caused by turbulent air, his stick motion contained a very pronounced oscillatory component, around 1.4 cycles per second, which at first was present even in periods of quiet atmosphere. The oscillation, or dither, phase-modulated to some extent by the main stick motion, persisted as long as the pilot anticipated pitch disturbances. When this dither was brought to the attention of the pilots at the end of the test series, they attributed it to the fact that they wanted to be certain of having continuous control of the aerodynamic situation, so as to be able to handle the next disturbance. According to the pilots, this type of control is particularly accentuated in situations where they are constantly under tension, e.g., formation flying and landing approaches.

Fig. 3 shows a partial recording of an experienced jet-interceptor pilot's test run. Channels 1 and 2 of the figure record air turbulence and the resultant airplane pitch angle, respectively. Channel 3 records the pilot's corrective motions. The oscillatory motion is obvious in this last channel, superimposed on what, without it, would have been a motion closely resembling that of channel 2.

The frequency of the oscillatory motion is of considerable interest in itself. In an ideal case, where the pilot attempts to follow a step input while the control loop is open, i.e., $\theta_a=0$ (see Fig. 2), and the step amplitude is much greater than the threshold level so that the threshold effect need not be considered, the two parameters characterizing the response will be the previously mentioned reaction time, τ, and the muscular-lag time constant, τ_0. The latter being the predominant kinesthetic factor.

When typical values for the two time constants, $\tau=0.25$ second and $\tau_0=0.125$ second,[3] are used and the pure gain factor is disregarded, a transfer function between input (stimulus) and response (control-force output of the arm) can be presented as the following equation

$$G(s)=\frac{\epsilon^{-0.25s}}{0.125s+1} \qquad (1)$$

where s is the Laplace operator.

* GEDA is a registered trade-mark of Goodyear Corporation.

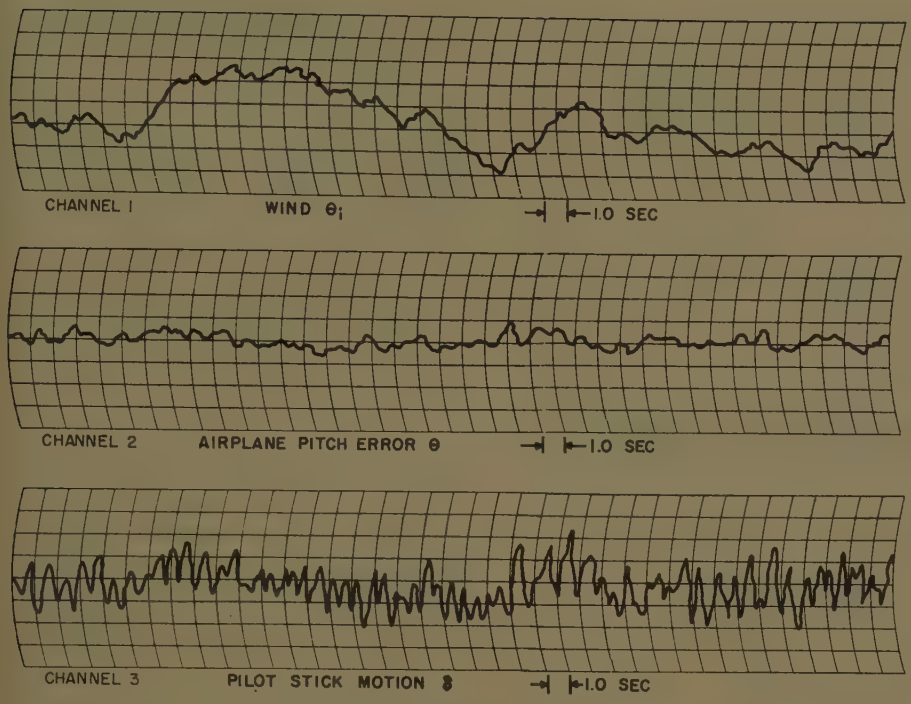


Fig. 3. Partial recording of pilot's test run

In the frequency domain, the reaction time factor, $\exp(-0.25\omega)$, becomes

$$\epsilon^{-j0.25\omega} = \cos(0.25\omega) - j\sin(0.25\omega) = 1/\underline{\phi = -0.25\omega} \quad (2)$$

which means that the phase shift, ϕ, is proportional to the frequency. Similarly, the muscular-lag factor, $(0.125s+1)^{-1}$, becomes

$$(1+j0.125\omega)^{-1} = \frac{1-j0.125\omega}{1+(0.125\omega)^2} = \frac{1}{\sqrt{1+(0.125\omega)^2}}\ \underline{/\phi_0 = -\tan^{-1}(0.125\omega)} \quad (3)$$

Therefore, the total phase shift is

$$\Phi = \phi + \phi_0 = -[0.25\omega + \tan^{-1}(0.125\omega)]$$

in radians. When $\omega = 2\pi f$ and $\Phi^\circ = 180/\pi\ \Phi$ are substituted, the function can be presented as

$$\Phi^\circ = 57.29[1.570f + \tan^{-1}(0.785f)] \quad (4)$$

The plot of Φ° versus frequency in cycles per second of Fig. 4 shows that the phase reversal, $\Phi^\circ = 180$, occurs at about $f = 1.5$ cycles per second which is practically the center of the observed dither frequency range.

Informative Feedback

In the study conducted, the presence of dither was unmistakably observed during the performance of jet pilots. Its cause can be looked for in the way high-speed aircraft have to be flown. Piloting depends on the knowledge of the ship's performance as influenced by its high speed, that is, on the knowledge of the performance characteristics of the plane-air system. In continuously acquiring this knowledge the pilot imparts a succession of small fast impulses to the control column. Though these do not deflect the airplane from its course, they are quite sufficient to give the pilot a kinesthetic sense of whether the plane is in danger of getting out of hand. This source of pilot awareness and its control effects are known in cybernetics as informative feedback.[4]

The relationship between feedback and control output permits a planned variation of what in servo terminology would be called the pilot's transfer function. The pilot superimposes a weak high-frequency input, δ_i, on the control signal δ_c, and takes from input θ a partial output, θ_f, of the same high frequency separated from the rest of θ through an appropriate filtering process; see Fig. 2. He explores the amplitude-phase relations of the high-frequency output component, θ_f, to the input δ_i, to obtain the performance characteristics of the plane-air system. What he learns allows him to modify his own transfer function for optimum control, very much in the manner of a self-adjusting servo.

The advantage of this type of feedback is that the resulting adjustments can give stability for every type of constant load within the limitations of both operator and machine. If the characteristics of the load change slowly enough in comparison with the changes of the original input and if the reading of the load condition is accurate, the system is prevented from going into instability. It is assumed, of course, that the load characteristics at high frequencies are the same as, or at least give a good indication of, the load characteristics at low frequencies.

The aspect of the plane-air system that chiefly interests the pilot of a high-speed airplane is the g force caused by changes in the velocity vector that occur when the stick is deflected or the throttle setting changed. A high g force endangers both the airplane and pilot. The dither observed in the performance of the jet-interceptor pilots who cooperated in this human-dynamics study at Goodyear Aircraft Corporation was apparently the informative feedback necessary to inform the pilot as to how many g's he could safely pull at any moment and, therefore, how great a stick deflection he could afford without danger.

From the flight-simulation results pilots could be placed into three categories: those who applied no dither, those who did, and those who did but learned not to after repeated runs. The subject for whom no informative feedback was apparent (see Fig. 5) had nothing but conventional aircraft experience. All the jet pilots carried the informative-feedback habit of their normal flight experience to mock-up flight, see Fig. 6. One of the jet pilots exhibited no tendency

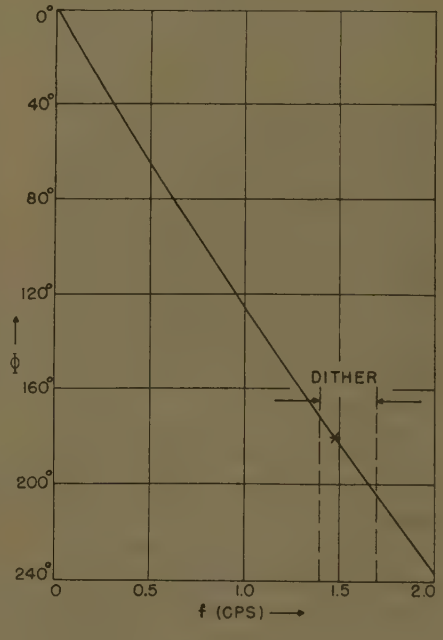


Fig. 4. Phase shift versus frequency in open-loop system

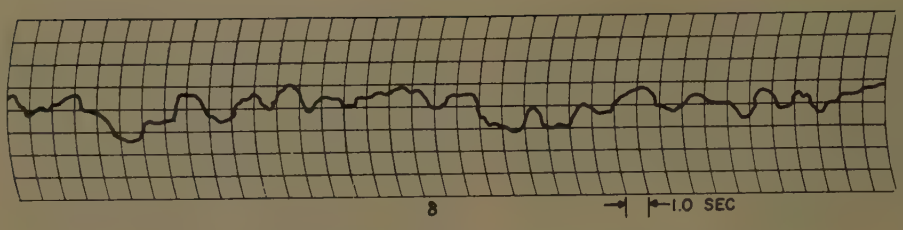


Fig. 5. Mock-up flight test record of conventional-aircraft pilot

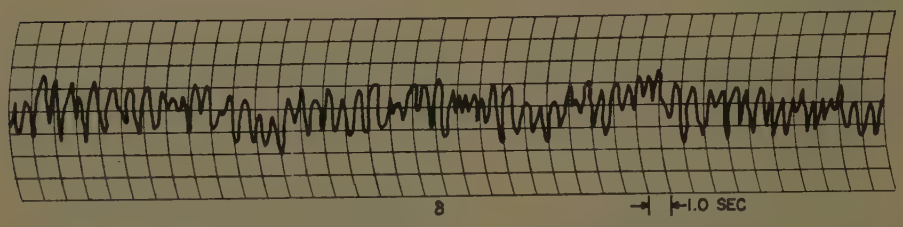


Fig. 6. Mock-up flight test record of jet-aircraft pilot

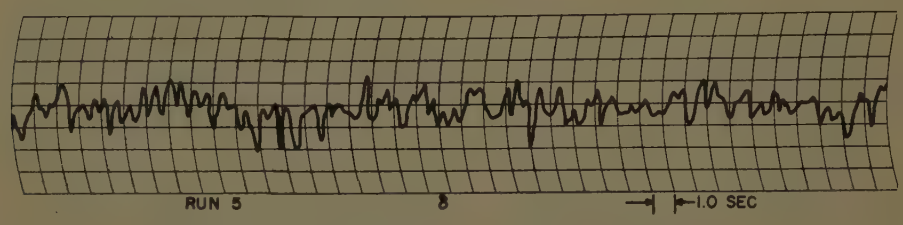


Fig. 7. Mock-up flight 5 showing presence of dither habit

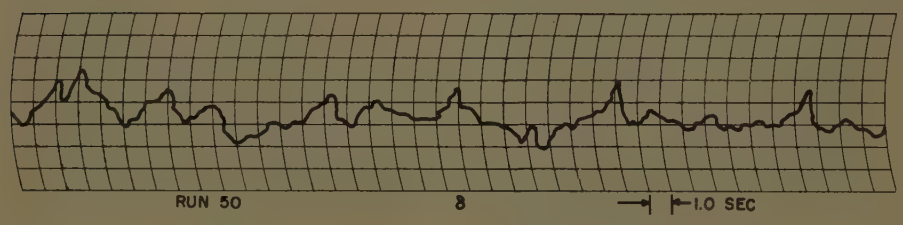


Fig. 8. Mock-up flight 50 showing loss of dither habit

of abandoning the habit, even after 80 10-minute trials. It is possible, of course, that even in his case dither would diminish or disappear entirely if the number of runs were sufficiently increased. Two other jet pilots started with dither motion but eventually learned that it was unnecessary for flying the mock-up where no actual g forces appear to endanger their personal safety, see Figs. 7 and 8.

Once the gradual dither extinction was ascertained, the runs of the next jet pilot were watched closely from that standpoint. A simulated pilot incorporating a dither oscillator and operating in parallel with the human pilot facilitated day-to-day measurements of the dither present in the subject's stick motion. Fig. 9 shows the gradual elimination of informative feedback as the subject's subconscious awareness of its uselessness increased. The experimental data represented in the figure could become the nucleus of a study in itself. The points on the plot show the dither amplitude for the last run of each successive day. It should be emphasized that these points were determined by visual comparison of the stick motions of human and dummy pilots operating in parallel and for the same θ input. No statistical approach was attempted.

Experimental Extinction of Dither

The gradual elimination of dither is a study to which psychological learning theory may be quantitatively applied. Since it is also a control-process phenomenon it is worth further analysis if for no other reason than that it may permit the interrelationship of two very different fields, namely, control engineering and psychology.

The decrease in the amplitude of informative feedback shown by the experimental points in Fig. 9 is a clear example of the behavioral principle known as "experimental extinction," which is the inverse of the habit-formation process. In the field of behavior dynamics it is assumed that an act that has been followed by a need reduction in a given situation will be reinforced to form a habit.

Actually, the acquisition of the dither habit can be related to the essential nature of the learning process. Just as inherited reaction tendencies consist of receptor-effector connections, so the process of learning consists in strengthening certain of these connections or establishing new connections.[5] It should be emphasized that physical pathways, which the term connection might represent to an electrical engineer, are not here implied. What is meant, ultimately, is a functional association between cause and effect. Neither psychology nor physiology has yet provided conclusive proofs as to what the physical representations of these associations are or where they are located in the brain.[6] Furthermore, while organic evolution rigorously tends to eliminate redundant organs or their functions, it often makes a single organ adaptable to a number of functions.[7] Thus, the nervous system is sufficiently plastic to allow (although not always) the transfer of learned reactions. For these reasons it becomes questionable whether such reactions can be sufficiently localized at present to permit accurate physical representation.

In controlling a high-speed aircraft, inherited tendencies cannot be considered dependable, since the human machine was never meant to fly by its own power nor to participate in sonic or supersonic motions with its own unaided controls. Therefore, a great many of the adaptive acts necessary for the undertaking of this biologically strange adventure must be learned. The use of informative feedback is quite patently such a learned act. It is the outcome of the process of habit formation.

In a jet-interceptor pilot the habit is probably acquired as follows. Confronted with the task of controlling the vehicle either for the first time or under tense conditions, the pilot does not possess as part of his evolutionary endowment the means of direct and instant "feel" of the aerodynamics involved. Yet he knows that these aerodynamics

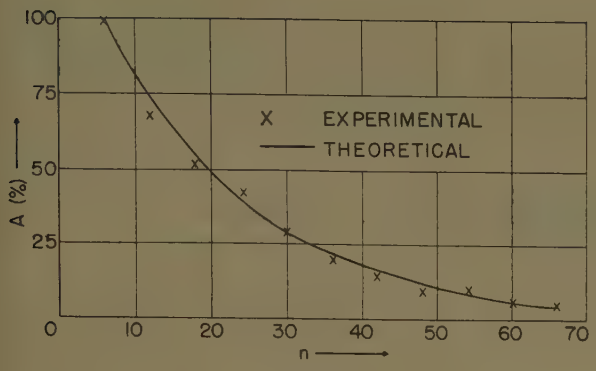


Fig. 9 (left). Elimination of dither by experimental extinction

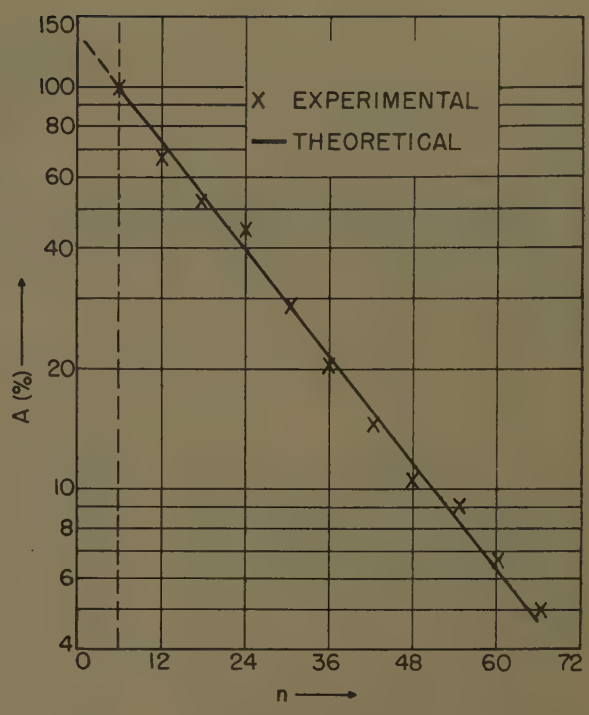


Fig. 11 (right). Logarithmic plot of experimental extinction

are varied continuously by the interaction of such factors as airplane characteristics, air speed, and air density and that his survival depends on his ability to alert himself to every change.

Apprehension of physical injury is the drive that evokes pilot action. Under the circumstances, dither is the action best suited to gain the knowledge required to allay his apprehension by providing continuous information on the plane-air system. Although the innate stimulus originating dither might be entirely casual at first, the dither reaction takes place in temporal contiguity with a stimulus energy that is followed by a diminution in the drive. That is, the output motion filtered and compared with the initial dither input provides the reassuring information being sought. The hypothesis of primary reinforcement suggests, therefore, that the tendency of the original stimulus to occur is likely to be incremented on subsequent occasions. Thus, a beneficial habit is gradually established and the threat to survival diminished accordingly.

In mock-up flying the informative feedback stimulus is not actually critical. The pilot, however, has no inner monitor to tell him in advance which stimuli, simple or complex, are associated with the critical causal factors of his dither reaction. Inevitably he reacts to the reinforced stimulus element associated with the dither.

Although the dither learning process is characterized by the setting-up and strengthening of specific receptor-effector connections, its effect is not limited to reinforcing these connections only. Connections are in fact mediated between a great many other receptor-effector processes. By no means are all the factors of any reaction situation critical to the extent that their presence is required in bringing about need reduction. In other words, the reaction involved in the original conditioning becomes connected with a considerable zone of stimuli other than, but adjacent to, the stimulus conventionally involved. This phenomenon, known as stimulus generalization, was certain to manifest itself in the experiment under discussion where the strong environmental similarities between actual and mock-up flying were bound to result in the overlapping of stimulus continua.

The mock-up contrasts with the plane-air system in one important respect: its transfer function remains fixed. Consequently, stimuli that previously evoked the informative feedback are divested of causal relationship to the critical factor in the reinforcement process. Although these stimuli continue to evoke the dither reaction, the reaction is not followed by reinforcement. Since the dither effort becomes wasteful and unadaptive, its gradual abandonment can be expected. The psychological axiom upon which this hypothesis is based is that a habit not followed by reinforcement can invariably be forced into experimental extinction.

A habit, functionally considered, is formed by receptor-effector connections set up by reinforcement. Likewise, the process of habit extinction consists of physiological subtraction of a series of discrete decrements, each decrement resulting from a distinct receptor-effector conjunction closely associated with the process.

Like that of any other habit, the strength of informative feedback cannot be determined quantitatively by direct observation. There are, however, two groups of measurable associated phenomena that make possible indirect measurement of feedback extinction: (1) the antecedent conditions that lead to the extinction and (2) the behavior persisting within the organism as determined by these antecedent conditions. The first group is the number, n, of preceding trials; the second is the dither amplitude, R, left in the stick motion. The fact that antecedent conditions persist within the system means that some sort of biological storage is involved.

The symbolism, $\dot{S} \rightarrow \dot{s} - - \rightarrow r \rightarrow R$, expresses the fact that a stimulus energy, $\dot{S}$, acting on a receptor, initiates an afferent receptor discharge, $\dot{s}$, that initiates an efferent discharge, r, which in turn enters the effector system and

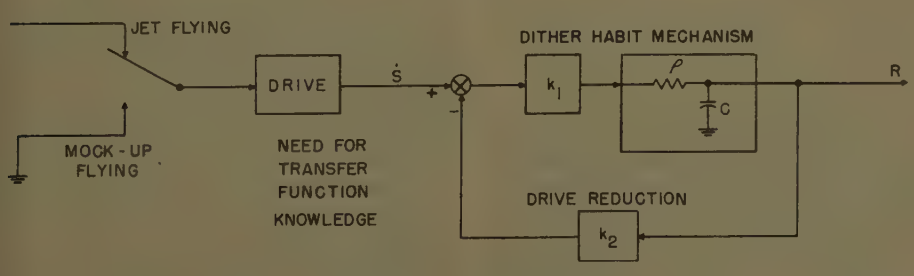


Fig. 10. Dynamics of habit

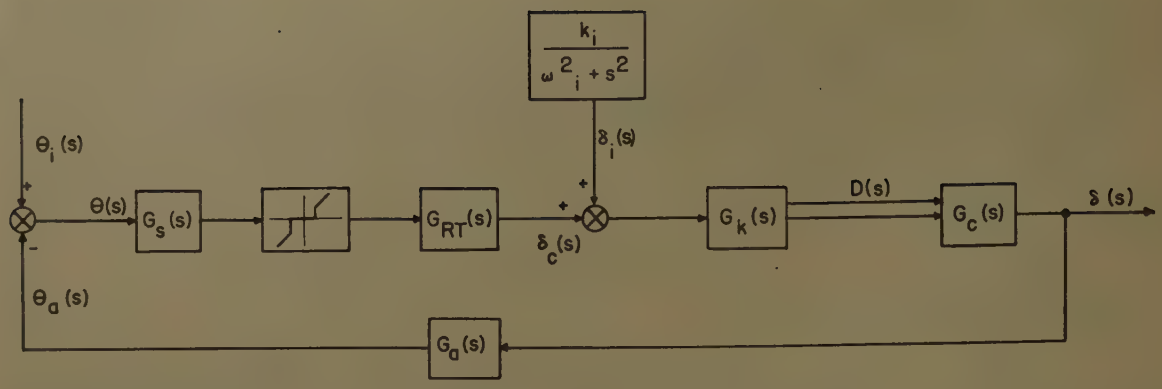


Fig. 12. Actua analog of closed-loop system

produces a reaction, R. It can reasonably be assumed that there is a one-to-one correspondence between $\dot{S}$ and $\dot{s}$ as well as between r and R. Hence, $\dot{S}\text{---}\rightarrow R$. Obviously the habit organization, ${}_{\dot{S}}H_R$, is represented by the broken-shaft arrow, and thus

$$\dot{S}\rightarrow {}_{\dot{S}}H_R\rightarrow R \tag{5}$$

This symbolic relation, although accepted in psychology, does not convey the notion of habit reinforcement. The foregoing discussion makes it plain that, from the standpoint of control engineering, habit formation and extinction are a closed-loop process. The habit mechanism is represented by the forward branch of the loop and the drive reduction by the feedback. Besides the aforementioned storage element, c, which is present in the mechanism, a dissipative element, ρ, should be included in view of the rather low efficiency characteristic exhibited by all biophysical functions. Consequently, in the way of a very rough approximation, the dynamics of the habit can be modeled as shown in Fig. 10, where the gains, k_1 and k_2, measure, respectively, the vigor with which the habit is acquired and the drive reduced. The response of the loop when the switch is in the JET FLYING position represents the formation of the dither habit. With jet training considered a step input to the drive complex of the pilot and with the transfer function of the foregoing oversimplified habit dynamics known from the diagram to be

$$R=\frac{k_1}{1+k_1k_2}\,\frac{1}{\dfrac{\rho c}{1+k_1k_2}\,s+1}\left(\dot{S}+R_0\,\frac{1+k_1k_2}{k_1}\right) \tag{6}$$

several facts can readily be observed. R_0 is the value of R at the beginning of the experiment. For instance, in the habit-formation stage the habit strength increases exponentially up to an upper physiological limit, beyond which no further increase is possible. The nature of the informative feedback suggests that there is a maximum value of dither amplitude, beyond which the airplane will begin to show an objectionable oscillatory motion. Clearly, this maximum amplitude will be the starting point, R_0, for the experimental extinction of the habit when the switch is thrown to the MOCK-UP FLYING position. Moreover, the greater k_1 and k_2 are, the greater the speed with which the dither is learned and abandoned. Factors such as the unconditioned stimulus and the amount of work involved in the habit mechanism influence the magnitude of the gains. It should be noticed that when the inverse Laplace transform of equation 6 is taken, the dither amplitude, R, is expressed as a function of the time, t, that has elapsed between the start of the experiment and the present. From a psychological standpoint, though, this time is effectively equivalent to the number, n, of trials performed during t. This is especially so when the trials are similar and of equal duration, τ_x. Hence, in the present experiment, $t\rightarrow\beta n\tau_x$. The parameter β depends on the design of the experiment. The greater the intertrial interval is, the smaller β is. Thus, the physical solution of equation 6 for mock-up flying, $\dot{S}=0$, becomes

$$R=R_0\epsilon^{-an} \tag{7}$$

where

$$a=\frac{(1+k_1k_2)\beta\tau_x}{\rho c}$$

Since the data of Fig. 9 were plotted in per cent of the sixth run, $n_0=6$, $R_0=R_6$, we find that the normalized dither amplitude, A as a function of the number of trials will be

$$A=\frac{R_n}{R_6}\,\epsilon^{-a(n-6)} \tag{8}$$

where

$$n\geqq 6$$

The derivation of equation 8, strictly speaking, should be based on the assumption that n increases continuously instead of by steps as it actually does, e.g., $n=6, 7, \ldots$. This is true during a particular run. The mathematical difficulty with the time interval between runs can be overcome if it is postulated that parameter β compensates for the effect the interval might otherwise have on the learning process. To calculate the constant, a, the experimental data were plotted on logarithmic paper (see Fig. 11) and, from the slope of the straight line best defined by the data, the value of a was derived as $a=0.0496$. Thus,

$$A=\epsilon^{-0.0496(n-6)} \tag{9}$$

A plot of equation 9 is given by the solid line of Fig. 9.

From equation 7 we find

$$(\Delta R)=-aR(\Delta n) \tag{10}$$

That is, the decrease in dither amplitude is proportional to both the number of antecedent trials, Δn, and the amplitude, R, at the beginning of Δn. This is a rather common empirical fact in many extinction experiments. Also significant is the fact that the model of habit dynamics, shown in Fig. 10, provides for the investigation of the relations between the response and each independently manipulatable environmental variable as well as the manner in which these variables interact. From an over-all point of view, equation 9 relates an actual instance of habit extinction to the theory that experimental psychologists have adduced on the basis of laboratory experiment.

Discussion

The results outlined indicate that analog simulation of psychological phenomena in real time could be a powerful tool in conducting controlled tests on the validity of models based on the biophysics of the human organism. Such a tool with an appropriate statistical design could eventually produce the quantitative expressions needed for objective description of the psycho-physiological activities of the human operator-expressions that are necessary to the study of control systems involving human participation. The limitations of the present analysis of informative feedback, though, are clearly recognized. The foundations upon which an approach of this type may be built are

none too firm at the present time, and of necessity a certain degree of speculative reasoning has been employed. The analysis will require ampler and more diversified experimentation.

Appendix

A pilot associates a certain stick force per g with the operating characteristics of the airplane that he is flying, i.e., speed, altitude, weight, and center of gravity. Therefore, the flight path attitude, γ, could be selected as the primary forcing variable, since the stick-load-factor increment is $V\dot{\gamma}/g$. In the simulated situation, however, the mock-up stick force was not affected by the flight path attitude since aerodynamic forces on control surfaces were absent. Consequently, the pitch attitude, θ, had to be selected as the primary variable in supplying the pilot with response cues. In other words, a function of θ constitutes the stimulus complex initiating the response.

In general terms, the transfer function for the stimulus block, $G_s(s)$, was based on the fact that the pilot is perceiving signals proportional to the actual airplane pitch, $a_0\theta$, and its first two derivatives, rate, $a_1 s\theta$, and acceleration, $a_2 s^2\theta$. Consequently,

$$G_s(s)=a_2s^2+a_1s+a_0$$

The control-column transfer function, $G_c(s)$, is the familiar second-order one of a system comprising inertia, I, compliance, $1/k$, and viscous damping, c:

$$G_c(s)=Is^2+cs+k$$

The airplane transfer function, $G_a(s)$, for the pitch-angle response, θ_a, to elevator motion, δ, is fairly simple since only one input is involved. For rapid tracking motions it can be further simplified by omitting the phugoid, or long-term pitch oscillation. Thus, for the particular jet plane simulated, this transfer function is

$$G_a(s)=\frac{K(\tau_1 s+1)}{s(\tau_2^2 s^2+2\zeta\tau_2 s+1)}$$

For the building blocks involved in the pilot simulator, no Laplace form for the threshold transfer function is possible since it involves a nonlinear operation. The pure time delay, however, which represents the physiological reaction time, τ, is known to have a transfer function of

$$G_{RT}(s)=\epsilon^{-\tau s}$$

The kinesthetic operation representative of the biomotor mechanism of arm and hand again involves elasticity, damping, and inertia in limb, muscle, and ligament. Its transfer function, therefore, can be given to a first approximation by a second-degree polynomial, $k'/(ms^2+ns+1)$, if open-loop performance is assumed. However, in a normal organism such performance will lead to ataxia, a cessation of communication between effector elements and the central nervous system. In this state the effectors, while remaining healthy, are no longer monitored. But such open-loop performance is not the case in a normal person. As a matter of fact, the proprioceptors throughout the muscles and joints of the body continuously monitor the effector organs. Their feedback signals are subject to a transmission time lag that can be approximated by an operation $k_0/(1s+1)$ in the feedback branch. Thus, the over-all kinesthetic transfer function becomes

$$G_k(s)=\frac{k'}{1+k_0k'}\times\frac{ls+1}{\frac{lm}{1+k'k_0}s^3+\frac{m+nl}{1+k'k_0}s^2+\frac{n+1}{1+k'k_0}s+1}$$

Finally, since virtually nothing is known about the search block, a 1.4-cycle-per-second oscillator, $k_4/(\omega_4^2+s^2)$, was substituted. The oscillation represents δ_4 as an independent input into the kinesthesis. On the other hand, the output of the $G_{RT}(s)$ block is $\delta_c(s)$. Fig. 12 is the actual analog used in this study. Control of the gain factor, k_4, by potentiometer adjustment permitted the dither amplitude of the analog to be matched to that of the human pilot, R. Thus the potentiometer readings provided a quantitative measure of R. The dither force resulting from δ_4 is illustrated in Figs. 1 and 2 as input D to the control stick.

References

1. Investigation of Control "Feel" Effects on the Dynamics of a Piloted Aircraft System. *Goodyear Engineering Report-6726*, Goodyear Aircraft Corporation, Akron, Ohio, 1955.

2. Application of GEDA to Human Dynamics Studies, R. J. Mead, N. D. Diamantides. *Aeronautical Electronics Digest*, Institute of Radio Engineers, New York, N. Y., 1955, p. 200.

3. Flight Controls Human Dynamic Response Studies. *Report AE-61-6*, Bureau of Aeronautics U. S. Navy, Washington, D. C., Nov. 1953.

4. Cybernetics (book), N. Wiener. John Wiley & Sons, Inc., New York, N. Y., 1955, p. 133.

5. Principles of Behavior (book), C. L. Hull. D. Appleton-Century Company, New York, N. Y., 1943, p. 258.

6. Significance of Information to Neurophysiology, J. A. V. Bates. *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, New York, N. Y., Feb. 1953, p. 140.

7. Possible Features of Brain Function and Their Imitation, W. Grey. *Ibid.*, p. 134.

8. The Kentucky Symposium on Learning Theory, Personality Theory and Clinical Research (book). "Interpretations of Learning Data and Some Recent Developments in Stimulus-Response Theory," K. W. Spence. John Wiley & Sons, Inc., pp. 3–4.

Flat Work-Coil Design

B. E. MATHEWS
NONMEMBER AIEE

IN RECENT YEARS, aerodynamic heating, or air-friction heating, has become one of the major problems in increasing aircraft velocities. Since little is known concerning the effects on aircraft structures heated in this fashion, experimental data must be obtained which will indicate necessary design procedure. It is desirable that the experiments be conducted under laboratory conditions to eliminate the necessity for flight, to give greater ease in instrumentation, and to give the greatest control of test conditions.

Since 1950, the University of Florida, under contract with the United States Air Force, has been conducting experiments to determine the feasibility of simulating aerodynamic heating of aircraft structures by induction heating. At present, a commercial 200-kw radio-frequency induction heating unit is being used as the primary power source. The results of these experiments will be applied to a test facility that will simulate both air friction heating and aerodynamic loading of aircraft structures. This paper is concerned with an analysis of the electrical characteristics of the work coil used to couple power into aircraft structures.

Paper 57-493, recommended by the AIEE Air Transportation Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE East Central and Middle Eastern District Meeting and Air Transportation Conference, Dayton, Ohio, March 7–9, 1957. Manuscript submitted March 5, 1956; made available for printing February 21, 1957.

B. E. Mathews is with the University of Florida, Gainesville, Fla.

The author wishes to thank Prof. Roger W. Sampson and Fred R. Sias of the Electrical Engineering staff at the University of Florida for their constant encouragement and for their invaluable assistance in preparing this manuscript.

Description of Work Coils

Using induction heating to simulate aerodynamic heating of aircraft structures imposes unique requirements on work coils and their design not usual in the more common uses of radio-frequency heating. Among the more important considerations are:

1. Surface materials are generally nonmagnetic and have low electrical resistivity. Therefore close coupling is necessary to give good efficiency.
2. To simulate effectively the spatial power

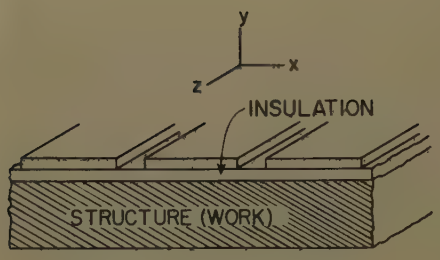


Fig. 1. Cross section of assembly

distributions resulting from aerodynamic heating, surface power should be as uniform as possible.

3. In a combined test using simulation of aerodynamic loading, much of the structure surface may be covered by pressure pads. Coils must fit under these pads.

4. Surface power should not change with warping of the structure.

Experience has indicated that the work coils consisting of wide, flat conductors lying close to the structure give good results. These coils have efficiencies which approach the maximum, produce fairly uniform power distributions, will fit beneath pressure pads, and when under pressure will follow changes in surface contour. In present testing, typical conductors may vary in width from one inch to 1/4 inch depending on the power density required and the current available. In most cases the coils are cut in one piece from copper sheet. Fluid cooling is not generally used and, since sufficient air cooling is difficult to obtain, the thermal capacity of the coil largely determines the amount of power which can be transferred to the structure without melting the coil. The thickness of the conductors is determined by a compromise between ability to follow changes in surface contour and thermal capacity. Generally, conductors with a thickness in the order of 0.050 inch are used.

Because of the voltage drop across the work coil, the coil is separated from the structure by sheets of insulation. Since close coupling is desired, the insulation thickness is kept to a minimum as determined by the voltage between the coil and the structure. A cross section perpendicular to the conductor axes of a complete assembly is represented in Fig. 1.

There are two general types of coils: the zig-zag and the spiral as shown in Fig. 2. The important difference between these is the relative direction of current flow in adjacent conductors and the resulting difference in the magnetic-field configuration. Applying a voltage across the zig-zag coil will cause a current which will be flowing in opposite directions in adjacent conductors. In the spiral coil the current will always be flowing in the same direction in adjacent conductors. These two types of coils have different electrical characteristics due to the difference in field configurations. The zig-zag coil and the spiral coil will, in general, take the form shown in Fig. 2 when heating one side of a structure. If two sides of a specimen are to be heated (or if the specimen is a figure of revolution), the spiral coil may take the form of a helix and encircle the specimen.

Method of Solution

Potential Analog

The work-coil characteristics (such as power density transferred, efficiency, or inductance) are determined by the distribution and magnitude of the fields produced by the work-coil currents. Therefore the work-coil analysis must begin with a description of these fields.

In general, if conductors are carrying current and are surrounded by an infinite, uniform dielectric and if the system is lossless, then the transverse electromagnetic (TEM) wave is propagated. In a TEM wave the electric and magnetic field are everywhere orthogonal and lie in the transverse plane (plane perpendicular to direction of propagation, or in this case, perpendicular to the axes of the conductors). The electric and magnetic fields are solutions of Laplace's equation and must satisfy certain boundary conditions at conductor surfaces.

In induction heating, the fields are not those of a true TEM wave because the conductors and dielectric are not perfect. However, a close approximation to a TEM wave exists when the following two conditions are met:[1]

1. Displacement currents in the conductors are negligible compared to the conduction currents.
2. The intrinsic impedance of the dielectric is much greater than the skin effect surface resistivity of the conductor.

Most metallic conductors and dielectrics in common use satisfy these conditions.

In practice, the region surrounding the conductors consists of a solid dielectric on which the current-carrying conductors are placed, and an air region or other dielectric above the conductors. Since this involves a composite dielectric, a pure TEM wave cannot be supported. Nevertheless, since in the flat coils with close coupling most of the field is concentrated in the dielectric between the work coil and work, the assumption of a single infinite dielectric introduces little error.[2]

The foregoing discussion leads to the conclusion that the field distribution produced by the work coils is, to a close approximation, the same as the distribution of a TEM wave. Also, the electric field produced by static electric charges is a solution of the same equation (Laplace's equation) and satisfies the same boundary conditions as the electric field in the TEM wave. Therefore the same electric field will be produced as in the TEM wave if the conductors are assumed to have static charges rather than to carry currents. In the construction of the electrostatic analogy the relative signs (positive or negative) of the charges correspond to the relative direction of the currents in the conductors, i.e., if adjacent conductors have currents which always flow in the same relative direction (spiral coil) the conductors will be given static charges of the same sign. If adjacent conductors have currents with opposite direction of current flow (zig-zag) the conductors will be given static charges of opposite signs. The electrostatic analogy will be used in this analysis to aid in visualizing the fields

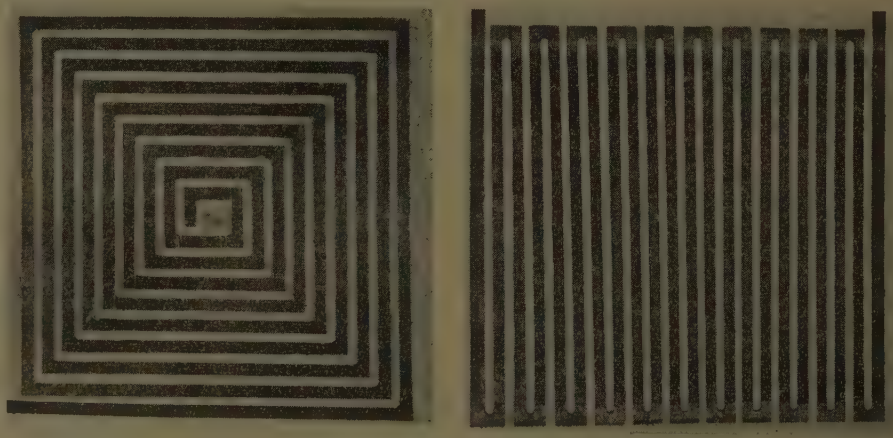


Fig. 2. Two general types of coils, zig-zag and spiral

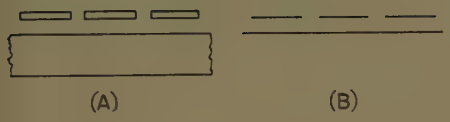


Fig. 3. Assumption of zero-thickness conductors

Fig. 4. Assuming wide strips

and because the analytic method used in the solution is more conveniently applied to the static case.

It should be understood that this electrostatic (or potential) analogy is used purely for convenience in describing the spatial distribution of the electric field in the TEM wave. In the actual case the fields are varying with time as the current varies, in this case sinusoidally. At any instant of time, however, this time-varying field has the same spatial distribution as the potential analogy. The potential analogy is used to describe the relative spatial field distribution but the magnitude and effect of this field will be based on the true time-varying conditions. The magnetic-field distribution follows from the electric field distribution since in a TEM wave the two fields are everywhere orthogonal and their magnitudes are related by the intrinsic impedance of the medium.

Now, consider an assembly cross section as that shown in Fig. 1. For the purpose of analysis it will be assumed that there are no field variations in the z-direction. This reduces the problem to a 2-dimensional case in the plane perpendicular to the direction of current flow. This assumes that the fields in the transverse plane are not being influenced by variations of the configuration in the z-direction such as corners where the conductor strips change direction, abrupt changes in the width of the strips, and variations in insulation thickness. The analysis on the basis of the 2-dimensional case will be the solution for all areas which have this cross section and which are not influenced by nearby discontinuities.

Solution Procedure in Electrostatic Potential Problem

The field distribution of the 2-dimensional electrostatic potential analog can be described with the aid of the theory of functions of a complex variable.[3] This method results in general equations that indicate broad design principles.

First a representation of the 2-dimensional configuration is drawn in the z-plane ($z=x+jy$). The conductors are then given appropriate potentials which will result in the desired field distribution as determined by the currents in the actual TEM wave. The next step is to determine the complex-potential function $P(z)$ which describes the field distribution This complex-potential function is an analytic function defined as

$$P(z)=\Phi(x,y)+j\Psi(x,y) \tag{1}$$

where curves of constant Φ are equipotential lines, and curves of constant Ψ are flux lines. Both Φ and Ψ satisfy the Cauchy-Riemann and Laplace equations.

The electric field in the z-plane is given by

$$E(z)=E_x+jE_y=-\left(\frac{\partial\Phi(x,y)}{\partial x}+j\frac{\partial\Phi(x,y)}{\partial y}\right) \tag{2}$$

The total derivative of the complex-potential function is defined as

$$\frac{dP(z)}{dz}=\frac{\partial\Phi(x,y)}{\partial x}-j\frac{\partial\Phi(x,y)}{\partial y} \tag{3}$$

Therefore from equations 2 and 3

$$E(z)=-\overline{\frac{dP(z)}{dz}} \tag{4}$$

where the bar indicates the complex conjugate. Returning to the actual TEM wave, the magnetic field $H(z)$ can be determined from the electric field since the two fields are orthogonal and related by the intrinsic impedance of the medium. The relationship is

$$H(z)=\pm\frac{j}{\eta}E(z)=\pm\frac{j}{\eta}\overline{\frac{dP(z)}{dz}} \tag{5}$$

where η is the intrinsic impedance. The sign of equation 5 depends on an arbitrary choice of the direction of power flow.

Equation 5 indicates that if the complex-potential function $P(z)$ is known, the magnetic- and electric-field distributions can be determined. For many simple configurations, the function $P(z)$ can be written by inspection. However, for complicated cases a transformation procedure can be used. Consider the configuration under study drawn in the z-plane. By the Swartz-Christoffel transformation, the configuration in the z-plane is transformed into the w-plane by a relation

$$z=F(w) \tag{6}$$

where $z=x+jy$, and $w=u+jv$. The resulting configuration is drawn in the w-plane, and the complex-potential function

$$P(w)=\Phi(u,v)+j\Psi(u,v) \tag{7}$$

which describes the fields in this plane can be determined by inspection. Due to the conformal transformation

$$\frac{dP(z)}{dz}=\frac{dP(w)/dw}{dz/dw} \tag{8}$$

Therefore from equations 5 and 8

$$H(z)=\pm\frac{j}{\eta}E(z)=\pm\frac{j}{\eta}\frac{dP(w)/dw}{dz/dw} \tag{9}$$

The electric and magnetic fields are now easily determined with equation 9 and the relationships of equations 6 and 7.

For the method of analysis as just outlined, it is convenient to make certain assumptions which simplify the actual configuration in the z-plane in order to reduce the mathematical complexity of the Swartz-Christoffel transformation. In general, simplification takes the form

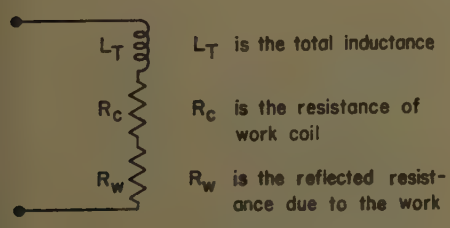


Fig. 5 (left). Representation of loaded coil

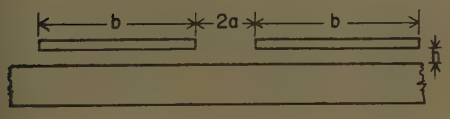
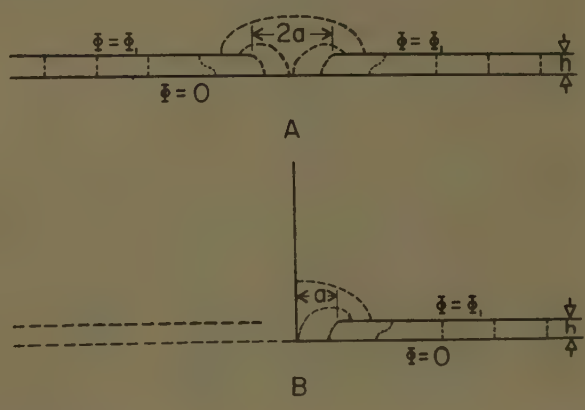


Fig. 6 (below, left). Assembly cross section

Fig. 7 (right). A—Simplified zig-zag cross section. B—Equivalent representation of A



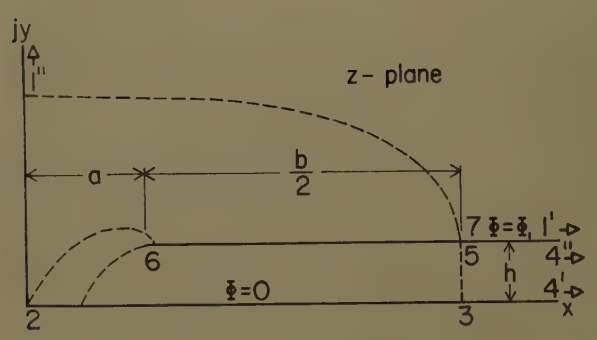
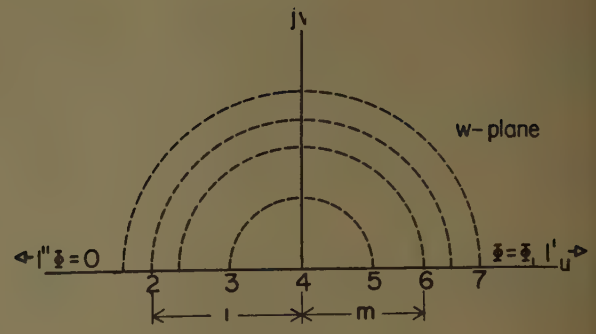


Fig. 8 (left). Cross section in z-plane

Fig. 9 (right). Cross section in w-plane



of reducing the number of "corners" in the z-plane.

First, the work coil and the work (structure) together are assumed to have zero thickness in constructing the potential analog. Fig. 3(B) represents the resulting cross section after this "zero-thickness" assumption is applied to the original cross section, represented in Fig. 3(A). This assumption is a good approximation because in the TEM wave, due to proximity effect and skin effect, the currents flow in very thin areas at the upper surface of the work and at the lower surface of the coil strips. It should be mentioned that after the field distribution has been described at the surface of the conductors with the use of the zero-thickness assumption in the analog, power calculations for the actual TEM wave are based on the assumption that all conductors are at least three skin depths thick. This is obviously necessary since power calculations based on zero-thickness conductors would be meaningless. These two seemingly inconsistent assumptions introduce little error since the field distribution based on zero-thickness conductors is approximately the same as the distribution based on small finite-thickness conductors. This is especially true for the flat-type, close-coupled coils.

The next simplification is based on the width of the work-coil strips being large compared to the spacing between the coil and the work (insulation spacing). Since this is true, the fields at the underside of the strip away from the edges will be uniform and therefore will have the same field distribution that would result if the strip were infinitely wide. Practically, this assumes that one edge does not affect the field distribution at the other edge or that one edge does not "see" the other edge. This assumption enables one to limit the section under investigation in the z-plane to include only two edges of adjacent strips while the other two edges are assumed to lie an infinite distance away. In Fig. 4 the broken lines border the area under investigation. The assumption is that in this area the field distribution in (B) is the same as the distribution in (A). The fields in other areas of (A) follow from symmetry.

Therefore, using the simplifying assumptions in the foregoing discussion, Fig. 4(B) represents the configuration in the z-plane. After the field distribution has been described, the various characteristics of the work-coil circuit can be determined.

Unless otherwise indicated, all expressions used are in the meter-kilogram-second practical system of units. In this system, the permeability of free space is $4\pi\times10^{-7}$ henrys per meter.

Power Calculations

If a sinusoidally time-varying magnetic field of rms value $H(x)$ exists at the surface of a plane conductor of infinite thickness, the time-average power density at any point on the surface is given by

$$P=H(x)^2R_s \text{ watts per meter}^2 \qquad (10)$$

where R_s is the surface resistivity defined as

$$R_s=\sqrt{\pi f\mu\rho} \text{ ohms} \qquad (11)$$

where

f = time-varying frequency, cycles per second
μ = permeability of conductor, henrys per meter
ρ = electrical resistivity of conductor, ohm-meters

The magnitude of the current in the conductor decreases exponentially with penetration into the conductor. The depth at which the current has decreased to $1/e$ (about 36.9%) of its value at the surface is termed one "skin depth" and is defined as

$$\delta=\sqrt{\frac{\rho}{\pi f\mu}} \text{ meters} \qquad (12)$$

Equation 10 can be used to approximate the power density in conductors of finite thickness with an error of less than 1.0% if the conductor is at least three skin depths thick. In the following analysis it will be assumed that the conductors (work and work-coil) meet this criterion, and equation 10 will be used in power calculations.

After equation 9 has been solved (it describes the magnetic field $H(z)$ everywhere in the z-plane), the expression for $H(z)$ along the conductor surface is substituted into equation 10 and the resulting equation integrated over the area under investigation. The resulting expression for the total power is then divided by the total area which gives the average power density over the area. This procedure can be applied to the work surface and the work-coil surface and will give the power coupled, the power loss, and the efficiency.

Impedance Calculations

The equivalent impedance of the loaded work coil may be represented by the series circuit shown in Fig. 5. Expressions for the resistive components R_e and $_w$ are obtained by dividing the power expressions by the square of the current.

The total inductance L_T may be divided into two parts: the inductance due to fields outside the conductors, referred to as the external inductance L_e, and the inductance due to the fields inside the conductors (work and work coil), referred to as the internal inductance L_i.

The external inductance L_e is calculated with the use of the equation which relates the instantaneous current in the conductor to the instantaneous stored energy in the magnetic field. This equation is[1]

$$\frac{1}{2}LI^2=\int_v\frac{\mu|H|^2}{2}\,dv \qquad (13)$$

The expression for H from equation 9 is substituted in equation 13, and the expression integrated over the area in the z-plane (assuming unit depth). This results in an equation for L_e.

The inductance due to the fields inside the conductors L_i can be easily calculated on the basis of semi-infinite conductors. For a semi-infinite plane conductor there exists an inductive reactance which is equal to the skin-effect resistance. In other words, the internal impedance has a phase angle of 45 degrees. Since the

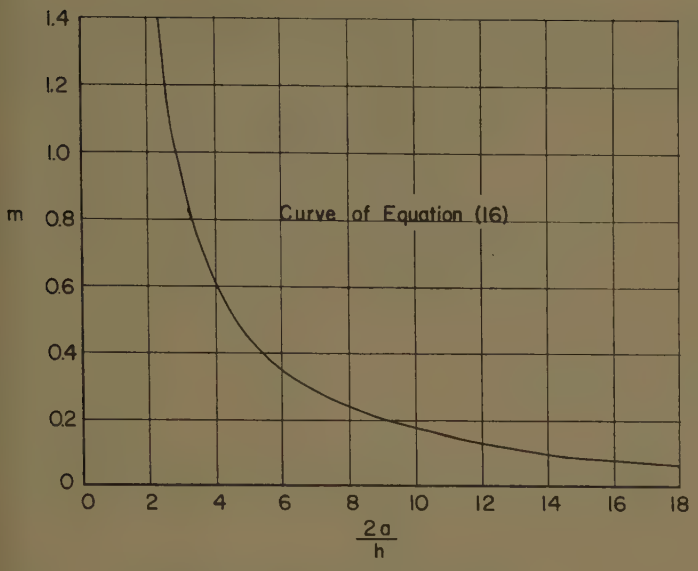


Fig. 10 (above). Curve of equation 16

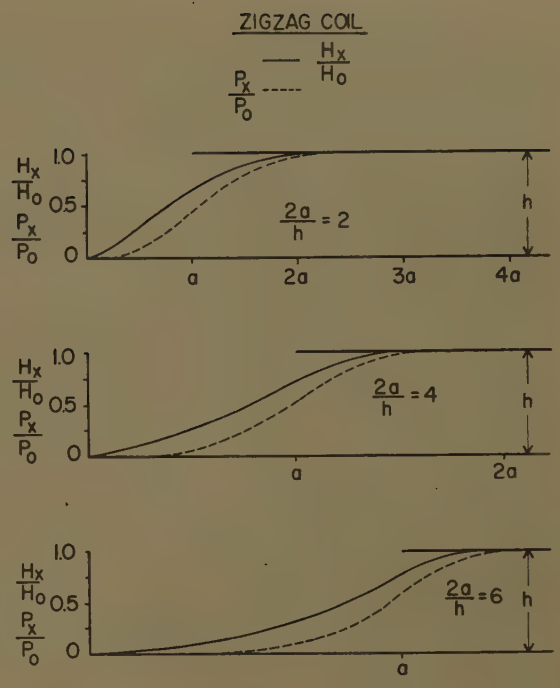


Fig. 11 (right). Field and power profile for zig-zag coil

work coil is assumed to be approximately a semi-infinite conductor (at least three skin depths thick) its internal inductive reactance will be equal in magnitude to the resistance of the work coil. Also it can be shown that, assuming the work is a semi-infinite plane conductor, the inductive reactance due to the fields in the work is equal in magnitude to the reflected resistance due to the work. Therefore, the magnitude of the reactance due to L_t is equal to the resistance "looking into" the loaded work coil.

The impedance of the loaded circuit is now fully defined. For many purposes it is convenient to define a quantity Q as

$$Q=\frac{\omega L_T}{R_c+R_w} \tag{14}$$

In the following analysis an equation for Q is derived in terms of the work-coil dimensions. Although the author has analyzed both types of work coils (zig-zag and spiral), the analysis of only the zig-zag coil is included in this paper.

Zig-Zag Work Coils

Construction of Potential Analog

Fig. 6 represents the cross section of the work-coil assembly and shows the notation which will be used to indicate the dimensions of the configuration. In the simplified cross section of the zig-zag coil, Fig. 7(A), the conductors are given appropriate potentials Φ to result in the desired field distribution. The work-coil strips are given equal and opposite potentials and the work is at zero potential. The dashed lines indicate probable electric-flux lines.

To make the problem even simpler consider the case of one work-coil conductor in a zero-potential corner; see Fig. 7(B). The field distributions in the upper-right-half planes of Fig. 7(A) and (B) are the same because of the image effect produced by the corner in Fig. 7(B). The field distribution in the upper-left-half plane follows from symmetry.

The configuration of Fig. 7(B) is now drawn in the z-plane (Fig. 8). Transforming the configuration by the Swartz-Christoffel transformation into the w-plane (Fig. 9) gives the transformation[3]

$$z=\frac{h}{\pi}\left[\frac{2}{m}\sqrt{w+1}+\ln\frac{\sqrt{w+1}+1}{\sqrt{w+1}-1}\right]+jh \tag{15}$$

where m is defined by the correspondence of point 6 in both planes and is given by

$$\frac{2a}{h}=\frac{2}{\pi}\left[\frac{2}{m}\sqrt{m+1}+\ln\frac{\sqrt{m+1}+1}{\sqrt{m+1}-1}\right] \tag{16}$$

Equation 16 is plotted in Fig. 10.

Note that the numbers in Figs. 8 and 9 represent points which correspond in both planes. Also corresponding flux lines are represented in the two planes.

The complex potential function in the w-plane is by observation:[3]

$$P(w)=\frac{-j}{\pi}\Phi_1 \ln w \tag{17}$$

Magnetic-Field Distribution

Differentiating equations 15 and 17 and substituting in equation 8 gives for the magnetic field

$$H(z)=\frac{\Phi_1}{\eta h}\frac{m\sqrt{w+1}}{(w-m)} \tag{18}$$

Although the magnitude of Φ_1 is purely arbitrary, it can be shown that $\Phi_1/\eta h$ is the magnitude of the constant (space-wise) magnetic field which appears between the work-coil and the work as x approaches infinity in the z-plane. Practically, however, this is the value of the field in the region of points 3 and 5. For convenience let

$$H_0=\frac{\Phi_1}{\eta h} \tag{19}$$

where H_0 is the magnitude of this constant magnetic field.

It is of interest to examine the fields at the surface of the work. In the w-plane, the work surface (point 2 to point 4, (Fig. 9) is along the u-axis from u equal to negative unity to u equal to zero. Therefore an expression for the surface field H_x can be obtained from equation 18 by letting w equal u and using the identity of equation 19. Equation 20 gives the ratio of H_x at any point to the maximum field H_0 as x approaches infinity.

$$\frac{H_x}{H_0}=\frac{m\sqrt{u+1}}{(u-m)} \tag{20}$$

The ratio H_x/H_0 can be plotted as a function of x (distance along work surface) with m as a parameter. (Fig. 9 has shown that m is a function of the ratio $2a/h$.) However, equation 20 expresses H_x/H_0 as a function of u, the distance along the conductor in the w-plane. The transformation equation 15 relates x to u since at the work surface $z=x$ and $w=-u$, but it is impractical to express u as a function of x from this equation. Therefore the best procedure is to pick values of u (from -1 to 0) and solve for the corresponding values of H_x/H_0 from equation

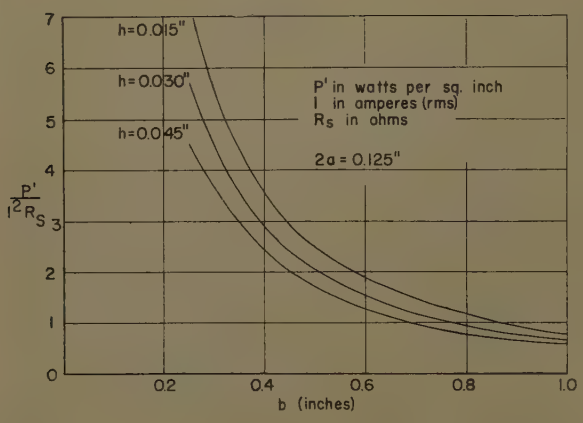


Fig. 12 (left). Coupled power density for zig-zag coil, curves of equation 34

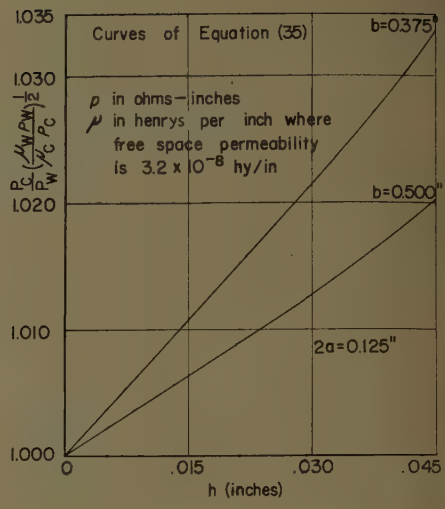


Fig. 13 (right). Zig-zag coil, curves of equation 35

20, and the corresponding values of x from equation 15. Then a curve of H_x/H_0 versus x can be plotted for each value of $2a/h$ (which determines m). In Fig. 11 curves are plotted for $2a/h$ equal to 2, 4, and 6 that give an understanding of how the field at the work surface varies as a function of x and $2a/h$.

Power Calculation

As indicated in equation 10, the power varies as the square of the magnetic field. Therefore, assuming constant surface resistivity, the ratio P_x/P_0 (the ratio of the power at any point on the surface to the constant power as x approaches infinity) varies as $(H_x/H_0)^2$. These curves also appear in Fig. 11.

To make calculations of the actual magnitude of power, the value of H_0 must be expressed in terms of the coil dimensions and the current flowing in the conductors. Consider the well-known equation

$$\oint H \cdot dl = I \tag{21}$$

which states that around any closed path the integral of the dot product of the magnetic-field intensity and the elemental distance at each point on the path is equal to the current enclosed inside the path. Now, consider half of a work-coil strip as represented in Fig. 8. Point 7 represents a point on the top surface of the strip at $x=a+b/2$, and 5 is a point on the lower surface at $x=a+b/2$. A closed path along the surface of this half strip would be followed by going from point 5 to point 6 and then to point 7 since the conductor has zero thickness. This same path in the w-plane is the straight line from point 5 to point 7. For this half-strip, equation 21 becomes

$$\int_5^7 H x dx = I_1 \tag{22}$$

where I_1 is the current in the half-strip. From equation 20

$$H_x = H_0 \frac{m\sqrt{u+1}}{(u-m)} \tag{23}$$

Differentiating equation 15, while letting $z=x+jh$ (position of strip in z-plane) and $w=u$ (position of strip in w—plane), results in

$$dx = \frac{h}{\pi m} \frac{(u-m)du}{u\sqrt{u+1}} \tag{24}$$

Substituting equations 23 and 24 into equation 22 gives

$$I_1 = \frac{H_0 h}{\pi} \int_5^7 \frac{du}{u} \tag{25}$$

Performing the integration in 25 results in

$$I_1 = \frac{H_0 h}{\pi} (\ln|u_7| - \ln|u_5|) \tag{26}$$

where $\ln|u_5|$ is the natural logarithm of the absolute value of u at point 5 and $\ln|u_7|$ is the natural logarithm of the absolute value of u at point 7.

Now, u_5 and u_7 must be expressed in terms of the configuration dimensions a and h. From correspondence of points 5 and 7 in the two planes as expressed in the transform of equation 15, u_5 and u_7 both must satisfy the equation

$$a+b/2 = \frac{h}{\pi}\left(\frac{2}{m}\sqrt{u+1} + \ln\left|\frac{\sqrt{u+1}+1}{\sqrt{u+1}-1}\right|\right) \tag{27}$$

For $b>>h$, u_5 is much smaller than unity so that

$$\sqrt{u_5+1} \approx 1 \tag{28}$$

Therefore, when expressing u_5, equation 27 may be written as

$$\ln\left|\frac{\sqrt{u_5+1}+1}{\sqrt{u_5+1}-1}\right| = \frac{\pi(a+b/2)}{h} - \frac{2}{m} \quad (b>>h) \tag{29}$$

Since u_5 is much smaller than unity, equation 29 can be simplified (by using the first two terms of the series expansion of $(u5+1)^{1/2}$) to

$$-\ln|u_5| = \frac{\pi(a+b/2)}{h} - \frac{2}{m} - \ln 4 \tag{30}$$

It can be shown that when $b \geqq 6h$ and $2a \leqq 0.6h$, the logarithm term in equation 27 is insignificant when expressing u_7. With these restrictions $\ln|u_7|$ is approximated by

$$\ln|u_7| = \ln\left\{\left[\frac{\pi m}{2h}(a+b/2)\right]^2 - 1\right\} \tag{31}$$

By trial it was discovered that the sum of equations 30 and 31 does not have a restriction on the value of $2a/h$ for a good approximation of $\ln|u_7| - \ln|u_5|$. Although there is a restriction on equation 31 of $2a<0.6h$, $\ln|u_7|$ becomes much smaller than $-\ln|u_5|$ at values of $2a/h$ where the error in equation 31 is significant. Therefore there is very little percent error in the sum when the approximations are used. With only the restriction $b \geqq 6h$, the sum of equations 30 and 31

$$\ln|u_7| - \ln|u_5| = \ln\left\{\left[\frac{\pi m}{h}(a+b/2)\right]^2 - 1\right\} + \frac{\pi(a+b/2)}{h} - \frac{2}{m} - \ln 4 \quad (b \geqq 6h) \tag{32}$$

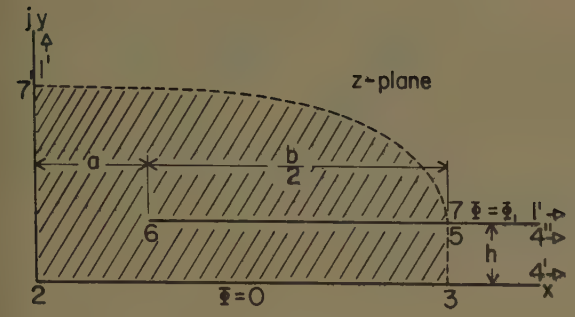
gives a good approximation of $\ln|u_7| - \ln|u_5|$ (equation 32).

Equation 26 may be rewritten

$$H_0 = \frac{I\pi}{2h} \frac{1}{(\ln|u_7| - \ln|u_5|)} \tag{33}$$

where $I=2I_1$, the total strip current, and $\ln|u_7| - \ln|u_5|$ is determined from equation 32. When equation 32 is used, it should be kept in mind that the approximations of equations 30 and 31 must be valid. If the approximations are not valid, u_5 and u_7 must be obtained from equation 27.

The average power density developed in the work is obtained by substituting equation 23 into equation 10, integrating equation 10 over the area from point 2 to point 3 (assuming unit conduction length), and then dividing by the area between points 2 and 3. The resulting equation



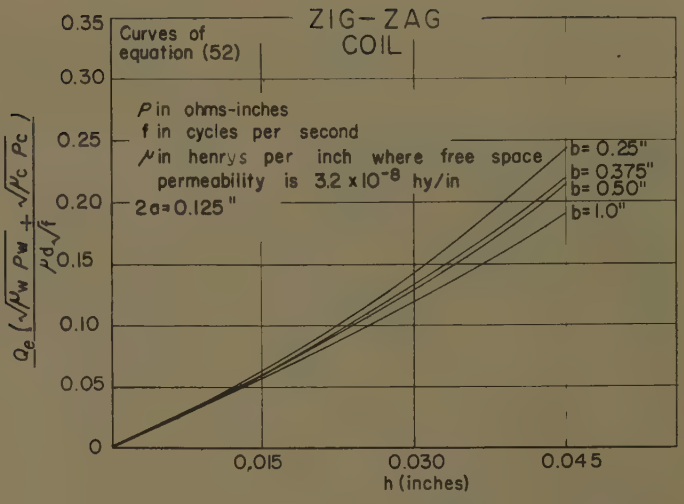


Fig. 14 (above, left). Area occupied by magnetic field from half-strip

Fig. 15 (left). Area occupied by magnetic field from half-strip

Fig. 16 (above). External Q for zig-zag coil

is as follows:

$$P' = \frac{R_s I^2 \pi}{4h^2(2a+b)(\ln|u_7| - \ln|u_5|)^2} \times [\pi(2a+b) - 2a\pi\sqrt{m+1} + 4h] \quad \text{watts per unit area} \quad (34)$$

where P' is the average power density developed in the work, R_s is the surface resistivity of the work, and I is the rms work-coil current. (Note: In deriving equation 34 it was again convenient to make the approximation of equation 28. The author considers that this approximation will be good for all practical cases.)

Fig. 12 shows the nature of equation 34. Here the width of the strip b is the independent variable with h as a parameter. The distance between strips was chosen to be 0.125 inch.

Efficiency

The power lost in the work coil can be obtained in the same manner as was the power in the work by integrating from point 5 to point 7. The expressions for the power coupled in the work and that lost in the coil may express the ratio

$$\frac{P_c}{P_w} = \left(\frac{\mu_c \rho_c}{\mu_w \rho_w}\right)^{1/2}\left(1+\frac{\alpha}{\beta}\right) \quad (35)$$

where

P_c = power lost in work coil
P_w = power coupled into the work
μ_c = permeability of work-coil material
μ_w = permeability of work material
ρ_c = electrical resistivity of the work-coil material
ρ_w = electrical resistivity of work material

and

$$\alpha = 4h\sqrt{m+1}\tanh^{-1}\left(\frac{4h\sqrt{m+1}}{\pi m(2a+b)}\right) \quad (36)$$

and

$$\beta = \pi(2a+b) - 2a\pi\sqrt{m+1} + 4h \quad (37)$$

Curves of equation 35 are plotted in Fig. 13 as a function of the spacing h with $2a = 0.125$ and b as the parameter.

The efficiency of the work coil is given by

$$\text{efficiency} = \frac{P_w}{P_w + P_c} = \frac{1}{1+\frac{P_c}{P_w}} \quad (38)$$

and may be calculated by using the results of equation 35.

The Q

As discussed in the section "Impedance Calculations," the external inductance is determined by the energy stored in the external magnetic field. Consider the z-plane represented in Fig. 14. All of the magnetic field produced by the half-strip is included in this area since the electric-field lines 3-5 and 7-7′ determine the outer limits of the field. In the w-plane (Fig. 15) this area lies between two semicircles.

As implied by equation 13, the external inductance is given (assuming unit conductor length) by

$$\frac{1}{2} L_e I^2 = \frac{\mu}{2}\iint |H(z)|^2 dx dy \quad (39)$$

Due to the conformality of mapping, the energy is also given by

$$\frac{1}{2} L_e I^2 = \frac{\mu}{2}\iint |H(w)|^2 du dv \quad (40)$$

Differentiating equation 17 as in equation 5 and using equation 19, the magnetic-field distribution in the w-plane is given by

$$H(w) = \frac{H_0 h}{\pi}\frac{1}{w} \quad (41)$$

Taking the absolute value of equation 41 and squaring it gives

$$|H(w)|^2 = \frac{H_0^2 h^2}{\pi^2}\frac{1}{u^2+v^2} \quad (42)$$

Substituting equation 42 into equation 40 and changing to polar co-ordinates results in (noting that total energy due to current in the entire strip is twice that due to the half-strip):

$$\frac{1}{2} L_e I^2 = \frac{\mu_d H_0^2 h^2}{\pi^2}\int_0^{\pi}\int_5^7 \frac{dr d\theta}{r} = \frac{\mu_d H_0^2 h^2}{\pi}(\ln|u_7| - \ln|u_5|) \quad (43)$$

where $\ln|u_7| - \ln|u_5|$ is defined by equation 32 and μ_d is the permeability of the dielectric material. Substituting equation 33 into equation 43 and solving for L_e, and dividing by the work surface area for one strip $(2a+b)$ gives

$$L_e' = \frac{\pi\mu_d}{2(\ln|u_7| - \ln|u_5|)(2a+b)} \quad \text{henrys per unit area} \quad (44)$$

where L_e' is the average external inductance per unit of work-surface area.

An expression for the resistance due to the work can be obtained from equation 34 by dividing by the square of the current, or

$$R_w' = \frac{P'}{I^2}, \text{ ohms per unit area} \quad (45)$$

where R_w' is the average resistance per

unit area of work surface due to the power developed in the work. From equation 38, the resistance due to the losses in the coil is

$$R_c' = R_w'(1/\text{eff (efficiency)} - 1), \quad \text{ohms per unit area} \tag{46}$$

As discussed in the section "Impedance Calculations," the reactance due to the internal inductance L_t is equal in magnitude to the total resistance of the circuit, or

$$\omega L_t' = R_w' + R_c' = R_w'/\text{eff}, \quad \text{ohms per unit area} \tag{47}$$

The Q of the circuit is

$$Q = \frac{\omega(L_e' + L_t')}{R_w' + R_c'} = \text{eff}\,\frac{\omega L_e'}{R_w'} + 1 \tag{48}$$

From equations 44, 45, 34, 35, and 38

$$Q = \left[\frac{4h^2\mu_d\sqrt{\pi f}(\ln|u_7| - \ln|u_5|)}{\beta(\sqrt{\mu_w\rho_w} + \sqrt{\mu_c\rho_c}) + \alpha\sqrt{\mu_c\rho_c}}\right] + 1 \tag{49}$$

Since in most cases (see Fig. 13) β is much larger than α, Q can be approximated as

$$Q = \left[\frac{4h^2\sqrt{\pi f}\mu_d(\ln|u_7| - \ln|u_5|)}{\beta(\sqrt{\mu_w\rho_w} + \sqrt{\mu_c\rho_c})}\right] + 1 \tag{50}$$

Equation 50 can be written

$$Q = Q_e + Q_t \tag{51}$$

where Q_t is the ratio of the internal impedance to the total resistance and is equal to unity, and Q_e is the ratio of the external reactance to the total resistance and is equal to

$$Q_e = \frac{4h^2\mu_d\sqrt{\pi f}(\ln|u_7| - \ln|u_5|)}{\beta(\sqrt{\mu_w\rho_w} + \sqrt{\mu_c\rho_c})} \tag{52}$$

Equation 51 is plotted in Fig. 16 as a function of h with $2a = 0.125$ and b as the parameter.

Application

Configuration Requirements

The equations derived in this analysis can be applied to any flat-type, close-coupled, zig-zag work coil whose configuration and application fit the description in the previous discussion. The most important configuration requirement is

$$b \geq 6h \tag{53}$$

In general, strips that satisfy equation 53 form a wide, close-coupled work coil which validates the wide-strip assumption and satisfies the requirements for the approximations in the analysis.

Accuracy

Several work coils have been constructed and their resistances and efficiencies measured with the use of a special low-frequency (around 300 kc) bridge. This measured data was, in general, within 10% of the calculated values for work coils of approximately 2 square feet in area. The greatest errors were encountered with small-area (one square foot) coils where the end effects (places where conductors change direction) constitute a large percentage of the total area. Very good correlation between measured values and theory was obtained with large-area (4 square feet) work coils.

The measured inductances were usually greater than the calculated values because of the added inductances of the leads. At present, there are insufficient experimental data to determine definitely the accuracy of the analytic expressions for Q.

Sample Calculation

For an example, consider the case where 2 square feet of magnesium alloy is to be heated with a power density of 500 watts per square inch with a generator used which is capable of producing 800 amperes and operating at a frequency of 400 kc. It will be assumed that the work material is at least three skin depths thick. A copper, zig-zag work coil will be used.

The following information is given:

$$\rho_w = 3.5\times10^{-6} \text{ ohm-inches (magnesium alloy)}$$
$$\rho_c = 6.7\times10^{-7} \text{ ohm-inches (copper)}$$
$$\mu_w = \mu_c = \mu_d = 3.2\times10^{-8} \text{ henrys per inch, meter-kilogram-seconds, nonmagnetic} \tag{54}$$

First, a value for the spacing between conductors ($2a$) is chosen. Past experience has indicated that $2a = 1/8$ inch (0.125 inch) is a good choice for most applications. Also, the necessary insulation thickness h is estimated considering the probable voltage across the coil and dielectric strength of the insulation. In this example, $h = 0.030$ inch is used.

The surface resistivity of this magnesium alloy at 400 kc is

$$R_s = (\pi f\mu\rho)^{1/2} = [\pi(4\times10^5)(3.2\times10^{-8})\times(3.5\times10^{-6})]^{1/2} = 3.7\times10^{-4} \text{ ohms} \tag{55}$$

In Fig. 12, the ordinate is

$$\frac{P'}{I^2R_s} = \frac{500}{(800)^2(3.7\times10^{-4})} = 2.1 \tag{56}$$

From Fig. 12, using the curve for $2a = 0.125$ inch and $h = 0.030$, the conductor width b for the ordinate 2.1 is

$$b \approx 0.50 \text{ inch} \tag{57}$$

Therefore the coil should be constructed with 0.50-inch conductors.

From Fig. 13 for this configuration

$$\frac{P_c}{P_w}\left(\frac{\mu_w\rho_w}{\mu_c\rho_c}\right)^{1/2} = 1.013 \tag{58}$$

Using the given values from equation 54

$$\frac{P_c}{P_w} = 1.013\left(\frac{6.7\times10^{-7}}{3.5\times10^{-6}}\right)^{1/2} = 0.44 \tag{59}$$

Then, from equation 38 the efficiency is

$$\text{eff} = \frac{1}{1+0.44} = 69\% \tag{60}$$

From Fig. 16

$$\frac{Q_e(\sqrt{\mu_w\rho_w} + \sqrt{\mu_c\rho_c})}{\mu_a\sqrt{f}} = 0.13 \tag{61}$$

Substituting the values from equation 54 into equation 61 gives

$$Q_e = 5.5 \tag{62}$$

From equation 51, the total Q is

$$Q = 5.5 + 1 = 6.5 \tag{63}$$

With this information, the characteristics of the work coil are completely defined. The reflected resistance due to the work is given by

$$R_w = \frac{(\text{power density})(\text{area})}{\text{current}^2} = 0.23 \text{ ohm} \tag{64}$$

The total resistance R_T is

$$R_T = \frac{R_w}{\text{eff}} = \frac{0.23}{0.69} = 0.33 \text{ ohm} \tag{65}$$

The total inductive reactance is

$$\omega L_T = R_T Q = 0.33(6.5) = 2.1 \text{ ohms}$$

The total inductance is

$$L_T = \frac{\omega L_T}{\omega} = \frac{2.1}{2\pi(4\times10^5)} = 0.84 \text{ millihenrys} \tag{66}$$

The total voltage across the work coil is

$$V = I|Z| = IR_T(1+Q^2)^{1/2} = 800(0.33)[1+(6.5)^2]^{1/2} = 1{,}700 \text{ volts} \tag{67}$$

References

1. Fields and Waves in Modern Radio (book), S. Ramo, J. R. Whinnery. John Wiley & Sons, Inc., New York, N. Y., 1944.

2. Simplified Theory of Microstrip Transmission Systems, F. Assadourian, E. Rimai. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 40, Dec. 1952.

3. Electromagnetic Fields (book), Ernst Weber. John Wiley & Sons, Inc., New York, N. Y., 1950.

The Role of Analog Computers in Propeller Control Design

ROY K. FRICK
NONMEMBER AIEE

IN THE PROCESS of designing and developing a dynamic system, provision should be made for analysis of the performance of the system while it is still in the design stage. One method of analysis is to write the equations of motion of each moving part, then combine these equations in such a manner that a means of solution is possible. This method is satisfactory if the system has few moving parts and thereby does not require many equations to describe its behavior. For a highly complex system this will prove to be very laborious by longhand methods and should be submitted to machine computation for solution. The motion of the parts of the system may be such that an equation describing the behavior of any one part may be difficult to write unless approximations are made. They may have to be directly simulated by a device which has the proper dynamic characteristics built into it.

The analog computer is especially suited to the analysis of control problems because it can compute the solution of differential equations and also simulate nonlinear phenomena directly. There are many forms of analog computers, some of which work by mechanical principles and others which work on electrical principles. The type of computer in widest use is that called the d-c electronic differential analyzer. It combines all the good features of the other types, such as flexibility, compactness, accuracy, and ease of maintenance.

The computer circuitry is wired in such a manner that the circuit equations are in the same form as the equations of the system under study. The voltages at various points in the circuit are analogous to the motions or positions of the parts of the system being studied. Because of these analogous voltage operations, the machine is called an analog computer. It is also called a simulator, although this term is generally applied to a special-purpose analog computer.[1]

Paper 57-501, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE East Central and Middle Eastern District Meeting and Air Transportation Conference, Dayton, Ohio, March 7–9, 1957. Manuscript submitted February 4, 1957; made available for printing March 14, 1957.

ROY K. FRICK is with the Wright Air Development Center, Dayton, Ohio.

Nomenclature

I = moment of inertia
N = rotational speed, rpm
ΔN = change in speed
β = propeller blade angle
$\Delta\beta$ = change in blade angle
ω_f = engine fuel flow
$\Delta\omega_f$ = change in fuel flow
Q_P = propeller load torque
ΔQ_P = change in load torque
Q_e = engine torque
ΔQ_e = change in engine torque
PLA = power lever angle
ΔPLA = change in PLA
T_4 = turbine inlet temperature
τ = time constant
(s) = Laplace operator
e_i = amplifier input voltage
e_0 = amplifier output voltage
Z_f = feedback impedance
Z_i = input impedance
i = instantaneous current
v = instantaneous voltage
X_c = capacitive reactance
J = advance ratio
V = air speed

Propeller Control Problem

A propeller is a device for converting mechanical shaft power into a propulsive force called "thrust." Thrust is a function of many variables, among these being the pitch of the propeller blades, the rotational speed of the propeller, and the air density. There are other variables, but these depend on the particular propeller being used. With the variable pitch propeller, a control system must be devised so that proper operating blade angle may be maintained. The type of control system most widely used governs the rotational speed of the engine by varying the propeller blade angle. By doing this, the control matches the propeller to the power of the engine.

With such a control system, the propeller governor must be able to sense an off-speed condition and then translate this into terms of blade angle. If an overspeed condition exists, the governor must increase the blade angle, which in turn will cause the load of the propeller on the engine to increase. This increase in load will cause the system to decelerate until such time as the rpm is again down to the desired value. If an overcorrection, or "undershoot" in this case, results, the propeller governor must reverse the process to bring the system back up to the desired value of rpm. The amount of succeeding corrections before the rpm settles out to its final value will depend on the characteristics of the propeller control and the flight condition of the aircraft.

In the design of propeller control systems, it is imperative to define the proper values of parameters, such as governor sensitivity, so that good performance will result. If a propeller governor is too sensitive to changes in rpm, excessive oscillating or "hunting" will result. If the governor and/or control system is too sluggish, excessive excursions in rpm may result. The sensitivity of the control must be a compromise such that all flight conditions of the aircraft are taken into account. For instance, a given change in propeller blade angle results in a change in propeller load torque which is larger at high air speeds than at low, and larger at low altitudes than it is at high altitudes.

The fuel control of the engine has certain dynamic properties and these must be recognized when a propeller control system is designed. Of course, the propeller control will have inherent characteristics of its own which the fuel control designer must appreciate. In any case the propeller control and the fuel control must be compatible. There must not be undesirable interactions or conditions under which the propeller control and fuel control have such unsatisfactory phase relationships that each control can oppose the corrections of the other.

The parameters which will determine proper compatibility of the propeller and engine control systems are such factors as time constants of servo systems within the propeller control, backlash in gears, true time lags of components, and dead zones of actuators.

To determine the proper values for the various parameters of the propeller control, the designer must make an analysis of the effect of each on the performance of the system. Methods of analysis generally involve approximations, but these are at a minimum if an analog computer is utilized for this purpose.

The gas turbine, or turbo-prop, engine is considered throughout this paper because this type of engine is in use in the latest propeller developments.

Discussion

Analog Computer Fundamentals

To understand the principles of analog computation, one must start with the fundamental concept of the transfer function. In Fig. 1 is shown the standard symbol of a transfer function. The transfer function G is said to operate on the input quantity e_i in such a way as to produce the output e_0. The relation is such that

$$e_0 = Ge_i$$

The transfer function concept is applied to an operational amplifier (Fig. 2), the basic component of a d-c electronic analog computer. The amplifier has input and output terminals which are accessible so that impedances may be wired into the circuitry of the amplifier. If an impedance is wired across the input

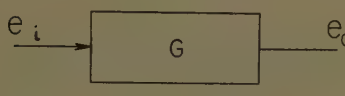


Fig. 1. Transfer function

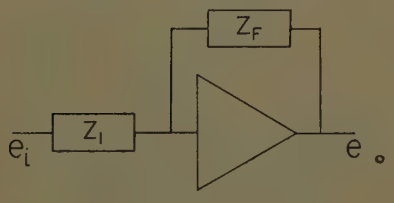


Fig. 2. Computer amplifier

and output terminals, it is referred to as a feedback impedance. If an impedance is wired only on the input of the amplifier, it is referred to as the input of the amplifier. The theory of the circuit design of the amplifier is such that the transfer function of the amplifier is equal in magnitude to the ratio of the feedback and input impedances and is opposite in sign. The transfer function G of the amplifier can be written

$$G = -\frac{Z_f}{Z_i} = \frac{e_0}{e_i}$$

where Z_f=feedback impedance, and Z_i=input impedance.

Depending on the types of impedances, in this case capacitive or resistive, an amplifier can serve as an integrator, a differentiator, or it may multiply by a constant. By wiring in various input impedances in parallel, an amplifier may serve as a summer, i.e., the various input quantities can be added. When the feedback impedance is capacitive and the input is resistive, the amplifier is an integrator. If the process is reversed, i.e., if the input is capacitive and the feedback resistive, the amplifier becomes a differentiator. By making both the feedback and input impedances resistive, the amplifier merely multiplies the input signal by the ratio of the feedback and input impedances. The theory of the transfer function of an operational amplifier is discussed in Appendix I. By utilizing the transfer function principal of the amplifier, one may proceed to set up a circuit which will solve linear differential equations with constant coefficients.

Basic Mathematics of Propeller Control System

A simplified engine and propeller system may be represented by a set of equations, some of which are differential equations. The equations given are representative of a propeller with a gas turbine engine.

The engine torque must accelerate the polar mass of the system and overcome the load of the propeller:

$$Q_e = 2\pi I \frac{d}{dt} N + Q_P \qquad (1)$$

or if only increments of torque and rpm are considered and equation 1 is rearranged

$$\Delta Q_e - \Delta Q_P = 2\pi I \frac{d}{dt} \Delta N \qquad (2)$$

The engine torque may be expressed as a function of fuel flow and rpm and the incremental equation for engine torque is

$$\Delta Q_e = \frac{\partial Q_e}{\partial N l_{\omega f}} \Delta N + \frac{\partial Q_e}{\partial \omega_f l_N} \Delta \omega_f \qquad (3)$$

Similarly the incremental propeller torque equation may be written with the knowledge that propeller torque is a function of blade angle and rpm.

$$\Delta Q_P = \frac{\partial Q_P}{\partial N l_\beta} \Delta N + \frac{\partial Q_P}{\partial \beta l_N} \Delta \beta \qquad (4)$$

Equations 2, 3, and 4 may be combined:

$$\left(\frac{\partial Q_e}{\partial N} - \frac{\partial Q_P}{\partial N}\right)_{l_{\beta\omega f}} \Delta N + \frac{\partial Q_e}{\partial \omega_f} \Delta \omega_f - \frac{\partial Q_P}{\partial \beta} \Delta \beta = 2\pi I \frac{d}{dt} \Delta N \qquad (5)$$

The following equation is generally used to express the operation of the propeller governor:

$$\Delta \beta = A \int^t \Delta N dt + \beta \Delta N \qquad (6)$$

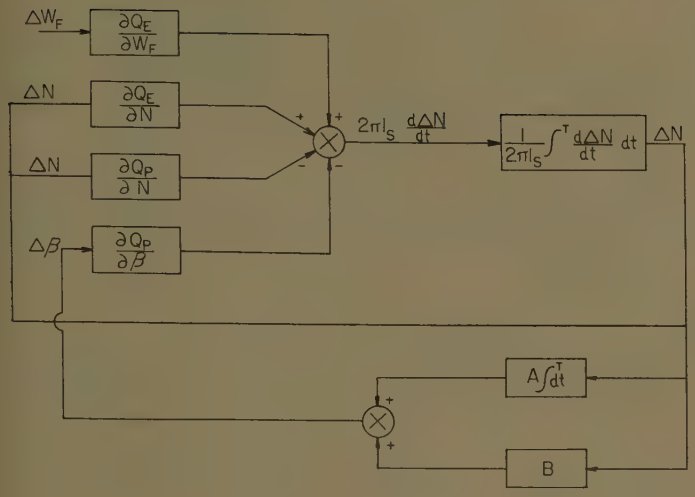
One extensively used practice is to draw a block diagram of the system before proceeding with the setup of the analog circuit. A block diagram of the system which has been described by the previous equations is presented in Fig. 3.

This block diagram is a collection of transfer functions connected together to form an over-all control loop. The advantage of the block diagram is that it shows the relationship of all the parameters and is a guide in setting up the analog circuit. The circuit after being set up is generally a duplicate of the block diagram. In Fig. 3, it is seen that each block is a transfer function and that these are determined from the coefficients of equation 5. The computer diagram for this block diagram is presented in Fig. 4. In Fig. 4 the plain triangular figures are summing amplifiers and the triangular figures with a double line are integrators. In many analog computers, the input and feedback impedances are internally wired and one has a choice of feedback and input combinations on a terminal board. When this is the case, symbols for summers and integrators are as those used in Fig. 4.

Fuel flow, rpm, and blade angle are multiplied by the appropriate partials in amplifier *1* which in turn produces a signal proportional to rate of change of rpm with a change in sign. Amplifier *2* integrates this signal and produces rpm increment with another change in sign. Rpm is integrated in amplifier *3* and changed in sign in amplifier *4*. These two outputs are added in amplifier *5* to produce positive blade angle increment. Since negative blade angle increment is needed for an input signal to amplifier *1*, amplifier *6* is needed to change the sign of the output of amplifier *5*. To complete the loop, rpm increment with the proper sign is fed from the output of amplifier *4* into the input of amplifier *1*. A check on the circuit for stability can be made by following the rule that there must be an odd number of amplifiers in any one loop. Reference to Fig. 4 indicates that this is true.

Fig. 4 shows that the fuel flow increment is the independent variable and that increments of blade angle and rpm are dependent variables. Other independent variables such as the change in rpm desired can be added to the circuit if appropriate for the particular system under study.

The computer circuit in Fig. 4 is a simulation of a propeller and engine system. It is a simplification, but one can use it to study the effects of the propeller and engine partials on the stability of the system at various flight



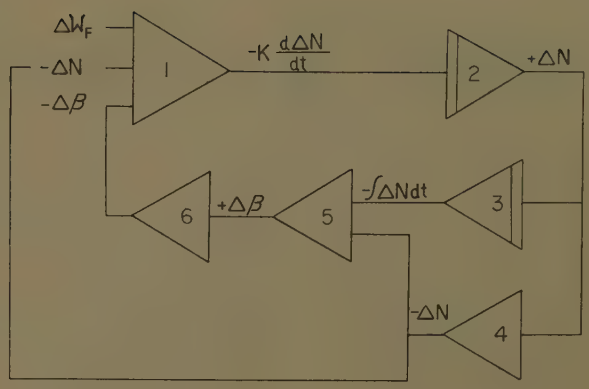


Fig. 3 (left). Linearized block diagram

Fig. 4 (above). Linearized computer circuit

conditions. Even though this circuit is not complex, the propeller designer can find satisfactory values for the partials of propeller torque to blade angle and rpm, and thereby influence blade characteristics. Satisfactory values for constants A and B can be found for equation 6 and this in turn will determine the proper governor design.

The circuit in Fig. 4, even though a simulation in probably its simplest form, is a far more effective tool in solving the equations than manpower alone. If equations 5 and 6 are combined, a third-order differential equation results and although not impossible to solve, would nevertheless consume quite a bit of effort. In addition, the solution would have to be tabulated and plotted and would not be quite as accurate as the computer solution.

Further Considerations

The simulation and analysis which has been explained to this point is called a linearized or perturbation type of study. This is so called because the equations are linear, i.e., constant coefficients, and the variables are treated as increments about a steady-state operating point. This type of study is valuable for evaluating the system for stability under various conditions. A high degree of accuracy is possible in studying the effect of the parameters on the stability of the system.

If extending the operating range of the variables is desired, a linearized system can no longer be considered. However, it may be desired to retain a fair degree of accuracy in studying the stability of the system, but to study the effect of various nonlinearities on general system operation may not be wanted. In addition, it may be desired to determine the transient characteristics to a higher degree of accuracy than is possible from a linearized computer setup. To proceed with this type of study a set of equations may be written which are based on available engine and propeller data. The equations will be empirical and will have no standard form from engine to engine and propeller to propeller. The following equations are illustrative.

The engine and propeller torque balance will still accelerate the system.

$$Q_e - Q_P = 2\pi I \frac{d}{dt} N \tag{7}$$

The engine torque may take the form

$$Q_e = (K_1 + K_2 T_4)N + K_3 T_4 + Q_{e_0} \tag{8}$$

and the turbine inlet temperature

$$T_4 = K_5 N + K_6 \omega_f + T_{4_0} \tag{9}$$

and the fuel flow

$$\omega_f = K_6 PLA + \omega_{f_0} \tag{10}$$

and the propeller torque

$$Q_P = K_7 \beta^2 + K_8 N + Q_{P_0} \tag{11}$$

and the propeller governor action

$$\beta = K_a \int^t N_e dt + K_{10} N_e + \beta_0 \tag{12}$$

The speed error is equal to the difference between the actual speed and the desired speed:

$$N_e = N - N_0 \tag{13}$$

and the desired speed is a function of power lever angle:

$$N_D = K_{11} PLA + N_{D_0} \tag{14}$$

The quantities with zero subscripts are initial conditions. Most makes of analog computers have special amplifiers incorporated which provide for setting and maintaining initial conditions.

The block diagram resulting from the previous equations is presented in Fig. 5. As before, the block diagram is the transfer function collection derived from a set of equations. The block diagram in Fig. 5 is similar in style to that in Fig. 3, but the circuit diagram in Fig. 6 is different in several ways from the diagram in Fig. 4. First, the diagram is much more extensive and, in an effort to simplify reading and setting up on the computer, the circuit schematic is broken up into segments. Labels such as encircled numbers are put on the various inputs and outputs; corresponding labels, in this case numbers, are joined together or are put at a common point when the circuit is being wired. One can see that if the labels were omitted on the drawing and solid lines were used for showing connections, the schematic would be harder to read and harder to follow when wiring the computer was attempted.

Second, the initial condition amplifiers are shown in the circuit. This is indicated by a triangular symbol which encloses the initials *IC*. Third, the symbol of a box with a diagonal is used to indicate a multiplier. A multiplier can multiply two or more variables or raise a variable to a power.

With this circuit, the engine and propeller system is very accurately simulated for purposes of studying transient operation over a reasonable degree of power excursions. It can also serve to study the stability to a reasonable degree of accuracy, though not as well as the linearized circuit.

Each amplifier which produces a signal in which one may be interested, say, rpm, may have the output utilized to drive a recording instrument such as an oscillograph. The various channels of the oscillograph will then record the variables as a function of time.

To simulate completely an engine-and-propeller system, the circuit in Fig. 6 would have to be further expanded. An expanded circuit is not shown, but the

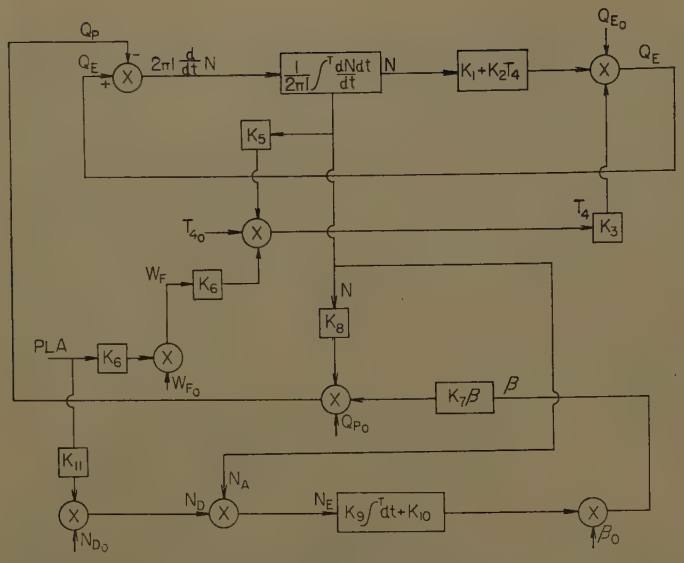


Fig. 5. Expanded block diagram

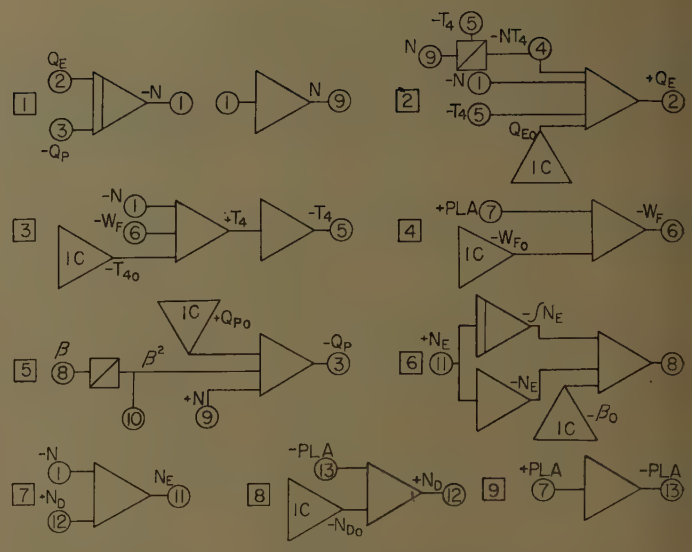


Fig. 6. Expanded computer circuit

additional equation and circuit complexities are discussed in succeeding paragraphs and in Appendix II. A lot depends on judgment as to how detailed a system should be simulated. It will depend on the factors in which one is interested and on the complexity of the system. These expanded operations will be necessary if, in the judgment of the engineer, the various second-order effects and nonlinearities are to be considered in evaluating the turboprop system performance.

Additional Complexities in Propeller

Equation 6 describes the propeller governor action which selects the proper blade angle to maintain the desired rpm. This equation is a close approximation for some types of propeller governors over the entire operating range. For other types of propeller governors, equation 6 will be valid only for perturbation-type studies where rpm and blade angle excursions are limited to small values. One feature which practically every propeller control possesses is the use of a servo system to translate the output motion of the governor into a blade angle motion. In this case, equation 6 is actually one which describes the desired blade angle instead of the true blade angle. The relation between the desired blade angle and the actual blade angle may be written as a first-order lag with a gain G so that

$$\beta = \frac{G}{1+\tau(s)}\,\beta_D \qquad (15)$$

This is assuming that such characteristics as dead band or backlash in gears is not significant.

In addition to this added complexity, the governor itself may employ servo systems of its own in generating the necessary terms in equation 6. Whenever these servos are employed, they may also be expressed by a first-order lag.[2,3] Equation 6 could become

$$\Delta\beta_D = \frac{A\int^t \Delta N dt}{1+\tau_2(s)} + \frac{B\Delta N}{1+\tau_3(s)} \qquad (16)$$

combining equation 16 with equation 15

$$(1+\tau_1(s))\Delta\beta = \frac{A\int^t \Delta N dt}{1+\tau_2(s)} + \frac{B\Delta N}{1+\tau_3(s)} \qquad (17)$$

As has been stated previously, equation 6 was valid for the entire operating range when some types of governor designs are considered. For other designs this is not true because the coefficients A and B are not constant. As an example, in many hydraulic propellers coefficients A and B are determined by the flow of fluid through a valve versus the valve displacement. The valve configuration may be a plunger which is enclosed by a sleeve with a circular hole. In equilibrium conditions, the plunger will completely cover the hole, but when displaced will uncover an area of the hole so that this exposed area is a function of the square of the displacement. The flow of fluid through the valve is directly proportional to the plunger displacement. This flow may be used to operate a servo system in generating coefficient B or may determine coefficient A.

Coefficients A and B in equation 6 are slopes of the curves depicting the relation of blade angle rate to rpm error and blade angle position to rpm error respectively. When A and B are nonlinear, it is simpler to write an empirical equation directly in terms of blade angle and rpm rather than trying to determine the slopes of the curves and in turn multiplying these by rpm error.

Equation 17 may turn out to be a power series in terms of rpm error and integral of rpm error.

$$(1+\tau_1(s))\Delta\beta = \frac{\int(A_1\Delta N^3+A_2\Delta N^2+A_3\Delta N)dt}{1+\tau_2(s)} + \frac{B_1\Delta N^3+B_2\Delta N^2+B_3\Delta N}{1+\tau_3(s)} \qquad (18)$$

The computer circuitry for this equation would be much more complex than that portion in Figs. 4 and 6, which simulate the governor action. If such things as dead band in governor or blade angle movement or backlash in gears are considered, the computer circuit must be further expanded to simulate these characteristics.

To raise a variable to a power such as that in equation 18, a multiplier is necessary. To generate a lag, a complex network is used for the feedback impedance of an amplifier. This feedback network consists of a capacitor and resistance in parallel. The development of this relation is left to the reader.

A circuit which can simulate dead zone is shown in Fig. 7. If the variable X is positive, but less than the absolute value of the constant A, amplifier *1* will have a positive output and the diode tube V_1 will not conduct. When the value of X exceeds the absolute value of A, amplifier *1* will have a negative

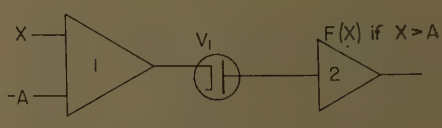


Fig. 7. Dead-band circuit

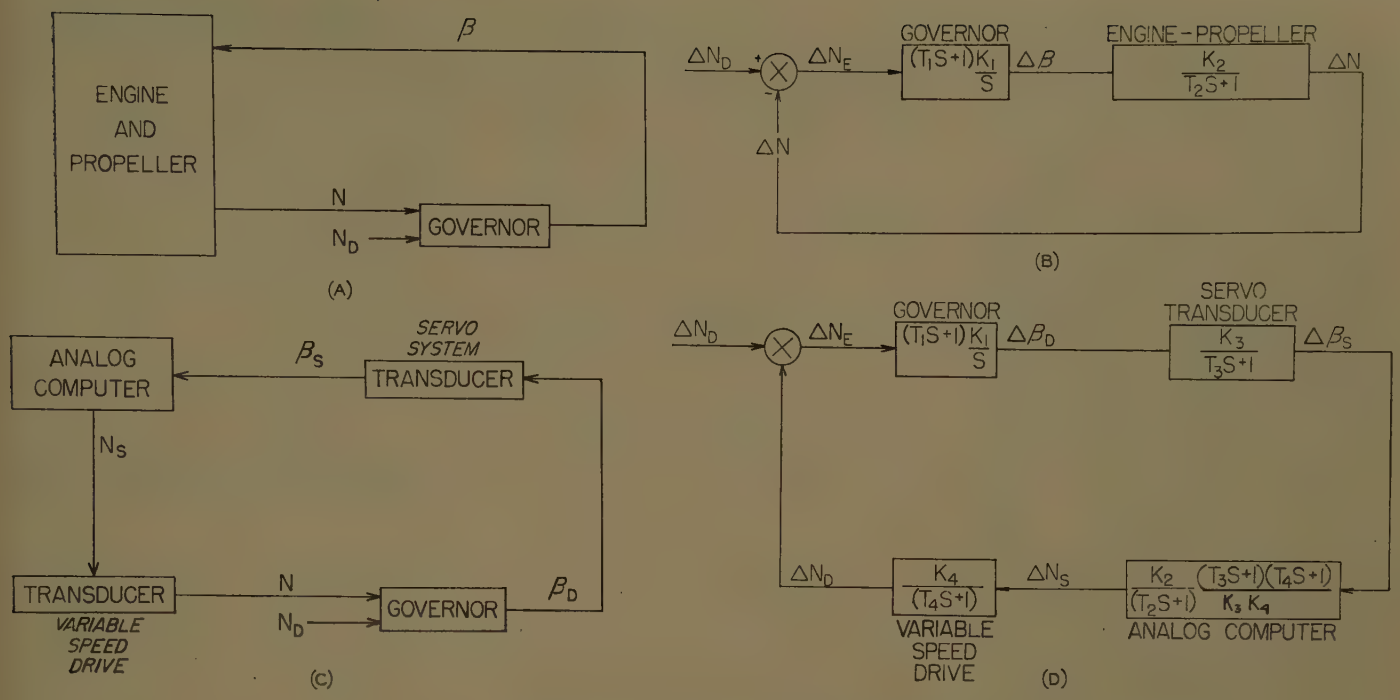


Fig. 8. Basic and equivalent system block diagrams

output and thereby allow tube V_1 to conduct. Amplifier 2 will then generate a function of X provided X exceeds A, the dead zone.

The various complexities such as lags and nonlinearities which have been mentioned can be added to the computer circuitry to simulate the system more accurately. The complexity is only limited by the equipment on hand and if the available equipment is limited, the system must be simplified. Valid approximations must be made for most problems, but made depending on the conditions of the problem. For instance, if to determine stability is desired, the equations may be linearized and characteristics such as dead band included. If it is desired to determine accurately the transient performance, empirical equations in the form of power series, may be written but such things as dead band approximated.

To evaluate properly the control performance of an engine and propeller system, a device such as an analog computer is a necessity. This is mainly because a highly nonlinear system such as an engine and propeller simply cannot be analyzed by longhand methods. Even if the equations are linearized, the analysis would possibly require the solution of a high-order differential equation which would be at best extremely long and laborious. Another method of analysis sometimes used for a system such as presented in equations 1 through 6 is called the step-by-step method. With this method values are calculated for each of the variables over a series of time increments such as 1/10 second. The initial torque unbalance between the engine and propeller is used to calculate the acceleration which in turn determines the rpm at a time interval later. The rpm is used to calculate the blade angle and in turn the engine and propeller torques. The new torque unbalance determines a new acceleration and the process is repeated. This repetition goes on until the rpm and blade angle settle out at a final value and the torque unbalance and thereby the rpm acceleration is equal to zero.

Use of Analog Computers in Component Testing

If a set of equations can represent a system such as an engine and propeller, it can be reasoned that one or more of these equations can be replaced by the actual component or subsystem it represents. To illustrate this, reference is made to Fig. 8. The analog computer essentially takes the place of the whole system with the exception of the component under test, in this case, the governor. Stated another way, the whole system is being simulated except the governor. If the computer works in terms of voltages, suitable transducers must be provided to translate the actual motions of the governor into voltage signals to the computer and to translate the voltage signals of the computer into suitable inputs to the governor. Most transducing systems have at least a first-order lag, so care must be exercised so that compensation is made for these lags in the over-all control loop. Fig. 8(A) represents in a general way the operation of a propeller control system. The governor accepts a signal which represents rpm and compares this to the desired rpm. The governor then translates this into an output which is proportional to the blade angle which is necessary to maintain the proper rpm. Whenever the blade angle changes, the load on the engine is changed which causes a change in rpm. Fig. 8(B) is a block diagram of an engine-propeller control system. The block representing the engine and propeller show that this part of the system acts as a time lag in relation to the blade angle and the rpm. This is actually a simplification, but for the purpose of this discussion, the assumption is valid. To provide good control performance, i.e., good transient and steady-state characteristics, the governor compensates for this lag by putting a lead term into the system.[3,4] Also to provide isochronous operation, the governor possesses an integrating term. The block representing the governor is equation 6 in the form of a transfer function. In Fig. 8(C), an attempt is made to build up a configuration which is dynamically similar to (A). As has been stated before, transducers introduce lags and compensations must be made for these lags. Fig. 8(D) shows the over-all loop where an analog computer

is used to compensate for the various lags and to simulate the dynamics of the engine and propeller. This setup is equivalent to Fig. 8(B).

Conclusions

Analog computers may be used in the propeller control design and development program for the study of dynamic performance, and they may be used in conjunction with component functional testing by simulating the flight characteristics of the engine and propeller.

Dynamic performance may be studied by solving the differential equations which describe the system performance. These equations may be solved by longhand methods provided the equations are of low order and are not highly nonlinear. Analog computation methods of analysis are superior to longhand methods because the time and manpower required are less and because fewer approximations are necessary. Solution of the equations results in establishment of criteria for satisfactory transient and steady-state control performance under all expected flight conditions.

The number of necessary differential and empirical equations in describing the system is determined by the number of unknown variables. Each of the equations describe the behavior of a particular subsystem or component. If one of these equations is deleted, a physical article which is described by this equation may be put into the simulation provided the proper transducers are available. When this is done, the overall system dynamics remain essentially the same, and control components can be subjected to simulated flight conditions. The analog computer can simulate the characteristics of the airplane, the engine, and the propeller blades, as well as the rest of the control system not under test.

A proper propeller control design will be possible only if adequate analysis and testing are accomplished. Analysis by use of analog computers results in the establishment of the proper values for all of the important parameters without having to resort to large-scale approximations, guesswork, or trial-and-error methods. Testing of components under simulated flight conditions will result in better ultimate performance and higher reliability.

Judicious use of analog computers in the design and development of propeller control system yields results in reduction of costs. These costs may be incurred from repeated engineering and design changes resulting from unsatisfactory control performance during the development program and in service.

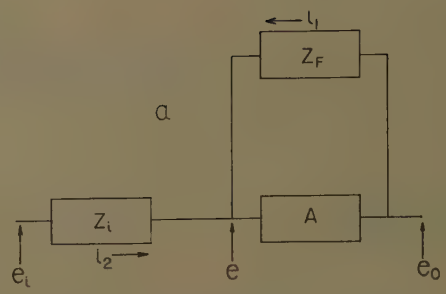


Fig. 9. Computer amplifier schematic

Appendix I

Fig. 9 is used to derive the transfer function of a computer amplifier. If it is assumed that there is no flow of grid current, the sum of the currents at point is zero, and it consists of the currents through the feedback and input impedances so that

$$i_1+i_2=0 \tag{19}$$

$$\text{but } i_2=\frac{e_i-e}{Z_{ii}} \tag{20}$$

$$\text{and } i_1=\frac{e_0-e}{Z_1} \tag{21}$$

The output voltage is equal to the gain A of the amplifier times the grid voltage e.

$$e_0=Ae$$

$$\text{and } e=\frac{e_0}{A} \tag{22}$$

combining (20), (21), and (22)

$$\left(e_0-\frac{e_0}{A}\right)\frac{1}{Z_1}=\left(e_i-\frac{e_0}{A}\right)\frac{1}{Z_2} \tag{23}$$

and

$$\frac{e_0}{Z_1}-\frac{1}{A}\left(\frac{e_0}{Z_1}+\frac{e_0}{Z_2}\right)=-\frac{e_i}{Z_2} \tag{24}$$

The gain A is generally a very large number on the order of 10^6 so that to an approximation

$$\frac{e_0}{Z_1}=-\frac{e_i}{Z_2}$$

or

$$\frac{e_0}{e_i}=-\frac{Z_1}{Z_2} \tag{25}$$

If the impedances are pure resistances, the transfer function of the amplifier is merely the ratio of the feedback and input resistances. If one of the impedances is a reactance and the other is a resistance, the amplifier will serve as an integrator or differentiator, depending on the combination. To illustrate this, consider the case of the feedback impedance of an amplifier being equal to a capacitive reactance and the input impedance equal to a resistance.

The charge on a capacitor is equal to the capacitance times the impressed voltage so that

$$q=Cv \tag{26}$$

$$\text{but } \frac{d}{dt}q=i$$

$$\text{or } v=\frac{1}{C}\int^t idt=\frac{1}{C(s)}\,i \tag{27}$$

$$\text{But } \frac{V}{i}=X_c$$

$$\text{so that } X_c=\frac{1}{C(s)} \tag{28}$$

where (s) is the Laplace operator.[2,3]

Substituting equation 28 in equation 25 gives

$$\frac{e_0}{e_i}=-\frac{1}{RC(s)} \tag{29}$$

where R is the input impedance or, restated

$$e_0=-\frac{1}{RC}\int^t e_i dt \tag{30}$$

Appendix II

The engine torque expressed by equations 3 and 8 is satisfactory for the intended purpose of these equations. However, in them the assumption is made that the engine torque which is delivered to the propeller shaft is instantaneous in response to the variables expressed. In other words, it is assumed that the engine torque follows the same relation to time as the fuel flow or rpm or turbine inlet temperature. For most propeller studies, it has been found that this assumption is valid and that a proper propeller control design will allow for variations in the engine torque dynamics. However, if one desires to simulate the engine more accurately, he must consider the dynamics of all the stations such as inlet duct, gear box, compressor, burner, turbine, and nozzle. Such things as pressures, air flows, temperatures, and specific heats would have to be considered and equations would have to be written to describe the various parameters in terms of known variables. A complete set of equations describing the thermal cycle of a turboprop engine is beyond the scope of this paper. One may consult any standard reference on gas turbine power plants to appreciate the complexity involved.[4,5] For the study of propeller control systems, it has been found that higher-order effects of the engine station are relatively insignificant. The dynamics of the fuel control should be considered, however, as experience has shown that interaction effects of the fuel control and propeller are very significant.

The fuel control of a typical turboprop engine is designed to maintain the fuel flow so that a nearly constant turbine inlet temperature exists for all flight conditions. This may be accomplished by a direct turbine inlet temperature controller or may be synthesized by sampling other quantities which the turbine inlet tem-

perature depends on. The fuel control must also protect the engine from overspeed, surge, or overtemperature conditions and must provide the engine with satisfactory response characteristics. The fuel control is made sensitive to such things as ambient temperature and compressor inlet or outlet pressure. This is done so that the fuel-air ratio will remain essentially constant for a given power setting and thereby cause the turbine inlet temperature to be indirectly controlled.

The nature of the fuel control is such that a simulation by an analog computer involves special devices such as switching circuits and limiting circuits. A detailed explanation of these circuits will not be presented here, but it may be seen that these circuits would be necessary to simulate fuel flow limiting at overtemperature and surge conditions, and fuel cutback whenever the rpm exceeds at certain value.[1]

The fuel control in sensing various parameters such as rpm, PLA, temperature, and compressor inlet pressure may employ servo loops to achieve a certain output. These servo loops probably will have transfer functions which are equivalent to a first-order lag.[2,3] If each of the parameters is characterized by a lag term where significant, the equations describing the fuel flow become very complex. From these considerations, it is seen that a true simulation of the fuel control is very complex and goes considerably beyond the circuit presented in Fig. 6. However, the circuit can be made highly representative of the actual fuel control operation, provided the necessary equipment to do the job is available.

The propeller torque relationship, if simulated accurately, can be the most complicated equation in the whole simulation and thereby result in circuitry which is quite elaborate. The propeller aerodynamic characteristics can be simulated to a high degree of accuracy and this is only limited by the availability of the necessary computing equipment.

The propeller torque can be expressed as

$$Q_P = c_p \rho N^2 D^5$$

where

c_p = coefficient of torque
ρ = air density
N = rpm
D = diameter

This equation can be rewritten as

$$c_p = \frac{Q}{\rho N^2 D^5}$$

For any given propeller, the diameter is a constant so that

$$c_p = K \frac{Q_P}{\rho N^2}$$

To determine c_p for a particular propeller, at a given set of conditions, the propeller may be calibrated for torque in terms of rpm and blade angle. In practice, the torque coefficient is determined for various blade angles, diameters, rpm, and forward airspeeds by using dynamically similar propeller models in wind tunnels. The torque coefficient is generally expressed as a function of blade angle and a term called "advance ratio."[5,6]

$$c_p = f(\beta_1 J)$$

where the symbols represent the following:

$J = V/ND$ = advance ratio
V = forward air speed

The torque coefficient may be presented as a family of curves with advance ratio and blade angle as independent variables. By a curve-fitting process, one may write an empirical equation for the torque coefficient. The resultant equation for propeller torque will involve at least a second power term in J, the advance ratio, and may involve higher orders of blade angle and rpm. This equation can be generated to a high degree of accuracy, but may require a considerable number of multipliers and amplifiers. A multiplier is needed every time a variable is raised to the second power or higher and whenever two terms are multiplied together.

References

1. Electronic Analogue Computers (book), G. A. and T. M. Korn. McGraw-Hill Book Company, Inc., New York, N. Y., 1952, pp. 1–21.

2. Engineering Cybernetics (book), H. S. Tsien. McGraw-Hill Book Company, Inc., 1954, pp. 7–23, 34–38, 63–66.

3. How to Specify the Performance of Closed Loop System, John E. Gibson. *Control Engineering*, New York, N. Y., Sept. 1956, pp. 122–29.

4. Aircraft Jet Powerplants (book), Franklin P. Durham. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1951, pp. 47–74, pp. 88–101.

5. A Dynamic Performance Computer for Gas Turbine Engines, V. L. Larrowe, M. M. Spencer. University of Michigan, Ann Arbor, Mich., Aug. 1955, pp. 6–45.

6. Fluid Mechanics (book), R. L. Daugherty, A. C. Ingersoll. McGraw-Hill Book Company, Inc., 1954, pp. 325–31.

Discussion

Irving A. Peltier (Wright Air Development Center, Dayton, Ohio): The author apparently sacrifices mathematical rigor for simplicity of presentation. Some examples follow:

He defines a transfer function G as the ratio between the output and input voltages of a system without defining these voltages adequately. If $G = e_0/e_i$ and e_0 and e_i are functions of time, then G is defined in the real time domain and its form is not particularly useful if the system contains time constants. A common method of expressing transfer functions in algebraic form is to assume that both the input and output voltages expressed as functions of time are Laplace transformable and then $G(s) = E_0(s)/E_i(s)$ where $E_0(s)$ and $E_i(s)$ are the Laplace transforms of the output and input voltages respectively. The transfer function $G(s)$ is in the complex frequency or s domain and is in a much more useful form since it describes the behavior of the output in response to many types of inputs readily.

Equation 27 is inconsistent.

$$v = \frac{1}{C}\int^t idt = \frac{1}{C(s)}\, i \qquad (27)$$

It should be written in two parts:

$$\text{If } v = \frac{1}{C}\int_0^t idt$$

then

$$V(s) = \frac{1}{Cs} I(s)$$

where $V(s)$ and $I(s)$ are the Laplace transforms of v and i respectively.

Equation 30 is correct if the initial value of e_0 is zero but the method of arriving at this equation is not convincing.

If the transfer function of a computer amplifier is derived with the use of Laplace transforms and the author's assumptions about grid current and output impedance, then

$$G(s) = \frac{E_0(s)}{E_i(s)} = -\frac{AZ_f(s)}{AZ_i(s) + Z_f(s) + Z_i(s)}$$

If A is very large, the approximation

$$G(s) = -\frac{Z_f(s)}{Z_i(s)}$$

is good over a large region containing values of s. However, if it is desired to differentiate e_i by making $Z_i(s) = 1/Cs$ and $Z_f(s) = R$, difficulty is encountered when e_i is changing rapidly with time. If it is desired to integrate e_i by making $Z_f(s) = 1/Cs$ and $Z_i(s) = R$, difficulty is encountered when e_i is changing slowly. The range of good approximation can be extended by proper choices of R and C but these values are very often limited by input and output impedance requirements.

The author says, "A check on the circuit for stability can be made by following the rule that there must be an odd number of amplifiers in any one loop." This neither constitutes a necessary condition nor a sufficient one for stability, even in linear systems.

It is reasonable to assume that handling nonlinear systems requires the services of an analog computer because mathematical methods are not available. However, the author should offer further proof of the analog computer's advantages over mathematical methods in the case of linear systems. It is conceivable that a mathematician might easily arrive at a single equation (transfer function) for a system. If the response of the system is too labor-consuming to plot, then he might, by modern synthesis methods, devise a passive network or semipassive network which has identically the same transfer function. This new system allows the original one to be studied for various inputs but does not allow it to be studied for changes in parameter or changes in the operation of basic components in the original system.

Economic Evaluation of Industrial Power-System Reliability

W. H. DICKINSON
MEMBER AIEE

IN THE DESIGN of an industrial plant, it is the function of the electrical engineer to determine what type of power system is required. This paper describes methods for the analysis of the relative reliability of different types of industrial power systems and the economic evaluation of the improvement in reliability which results with alternative types of systems. From this, the incremental investment that can be justified can be determined.

The paper covers four main points. First, a technique for the economic analysis of alternatives involving incremental investments is developed. The technique is a simplification of the discounted-cash-flow method of comparing alternatives. Second, probability and its application to power-system reliability problems is discussed. Third, these probability principles are applied to several typical industrial power systems. And last, a typical example is worked out. In this example, the incremental investment that is justified to increase reliability is evaluated.

Nomenclature

I = incremental investment
E = gross earnings (before income taxes, depreciation, and operating costs proportional to investment) resulting from I, expressed as uniform equivalent annual amount when non-uniform, that is,

$$E = a_n \sum_{n=1}^{n=n} \frac{E_n}{s_n}$$

R = investment charge factor
G = gross revenues or sales resulting from I
X = operating expenses (all costs not proportional to I)
C = cash flow or income (net income plus depreciation)
N = annual net income (after income taxes)
D = depreciation permitted by income tax authorities (straight line, or levelized or uniform equivalent value of sum-of-digits or declining balances), $D = j_n I$, per unit values
d_n = depreciation rate by sinking fund method, per unit value $= A_n/F_n$

$$= \frac{r}{(1+r)^n - 1} = \frac{r}{s_n - 1}$$

r = return rate (after taxes) per unit value, which is required on I
e = expense rate, per unit value, including all expense items proportional to I not included with X, usually includes maintenance and state taxes
t = income tax rate, per unit value
a_n = capital recovery factor $= A_n/P_1 = r + d_n$, per unit value
s_n = compound return rate factor $= F_n/P_1 = (1+r)^n$, per unit value
P_1 = present value of a future sum or sums of money at beginning of year one
F_n = value of a sum or sums of money in the future at the end of year n
n = number of years
A_n = annual amounts of money at the end of each year for n years, having a present value of P_1 at the beginning of the first year, and a future value of F_n at end of the nth year
$j_n = D/I$
p = probability of supply, circuit, or element not being on forced outage
f = probable annual equipment or supply failure rate causing a forced outage
S = shutdown time or forced outage hours per failure
q = probability of supply, circuit, or element being on forced outage

Technique for Economic Evaluation

In comparing power-system alternatives, the basic principle should be to determine the effect of each on the plant's cash flows. Cash flow is defined as net income plus depreciation, that is, depreciation is added to the profits after income taxes. Depreciation is included with the cash receipts or flow, since it is not an out-of-pocket expenditure. The discounted-cash-flow method of comparing alternatives follows this principle. It is believed to be superior to other methods which do not.[1-3] The discounted-cash-flow method gives a more realistic economic return irrespective of the time pattern of the earnings or the timing of investments. The differences between dollars today and dollars in the future are recognized.

Paper 57-639, recommended by the AIEE Industrial Power Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24-28, 1957. Manuscript submitted March 25, 1957; made available for printing April 25, 1957.

W. H. DICKINSON is with the Esso Research and Engineering Company, Linden, N. J.

The author wishes to acknowledge the assistance of his associates in the preparation of this paper, particularly that of E. J. Winsor, R. J. Knopf, and C. L. Wilkinson.

In the discounted-cash-flow method, the return rate is found that discounts the cash flow or receipts of the proposed alternative design to zero, counting all investments as negative receipts. This is equivalent to saying that the present worth of all the cash receipts is equal to the present worth of all the investments. The present worth of a sum of money at the end of year n is the sum divided by the factor s_n. The present worth of a series of n equal sums at the end of each year is found by dividing the sum by a_n. By multiplying the present value of a sum by a_n the uniform equivalent annual sums at the end of each of n years are found. For an explanation of any symbols not explained in the text, reference should be made to the nomenclature. In it, also, some of the financial mathematics necessary to deal with present and future sums of money are given. A disadvantage of the discounted-cash-flow method is that its solution requires cut-and-try methods. In order to obtain the advantages of the discounted-cash-flow method, but to avoid the complexity of its cut-and-try solution, a simplified technique has been developed.

Justifiable Investment

This technique uses the following equation, the development of which is given in Appendix I.

$$I = E/R \tag{1}$$

This is not Ohm's law. But it is somewhat analogous. It might be termed a financial Ohm's law. The investment, I, that can be justified is proportional to E, gross earnings; that is, the earnings before deduction of income taxes, depreciation, maintenance, state taxes, or any other expense that is proportional to investment. Similarly, I is inversely proportional to R. R might be termed an investment charge factor. It differs from a fixed charge factor in that it includes the required rate of return before income taxes.

The value of R in equation 1 is

$$R = \frac{r}{1-t} + e + \frac{d_n - j_n t}{1-t} \tag{2}$$

$$= \frac{a_n - j_n t}{1-t} + e \tag{3}$$

If, in equation 1, I and E are known, R may be determined. In the following equation, this value is substituted into equation 3. From which

$$a_n = (R-e)(1-t) + j_n t \tag{4}$$

By reference to tables, such as Table I, the return rate r may be found, if a_n is known. It also may be found by trial and error.

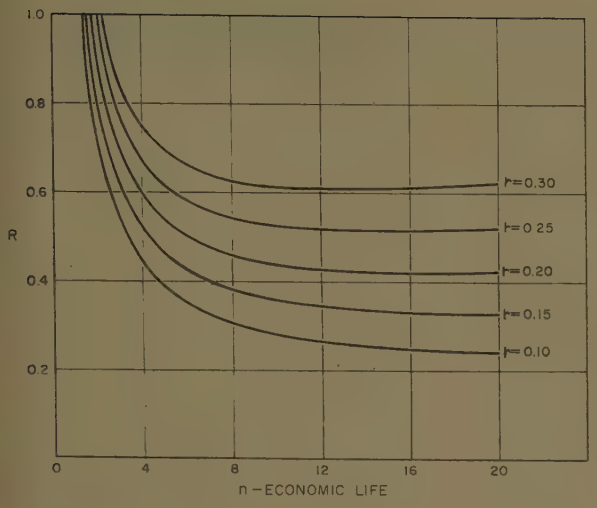
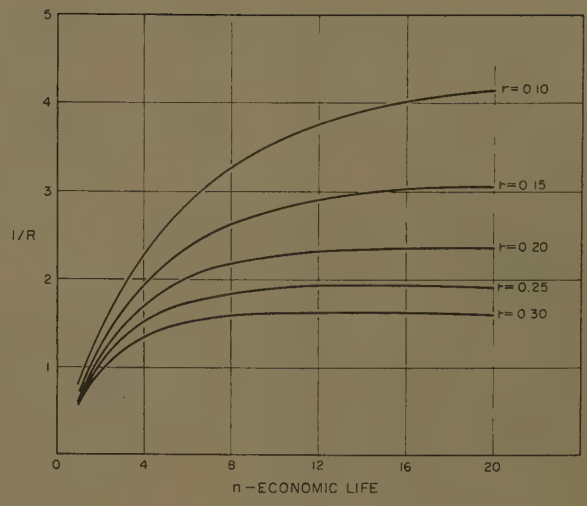


Fig. 1 (left). Values of R for e=0.05, t=0.52, $j_n=1/n$

Fig. 2 (right). Values of 1/R for e=0.05, t=0.52, $j_n=1/n$, investment justified per dollar of gross earnings, E



In the use of equation 1, it is intended that the minimum acceptable return rate r be first determined. Knowing this and the other factors in equation 2, R can be calculated. After E is determined, I, the maximum justifiable investment, can be calculated.

Values of R have been calculated for various return rates and plotted against n, economic life, for a typical case when $e=0.05$, $t=0.52$, and $j_n=1/n$. These curves are shown in Fig. 1. Similarly, the inverse of this, values of $1/R$ are plotted against n, as shown in Fig. 2. This gives directly the investment justified per dollar of gross earnings, E, by letting E equal \$1 in equation 1 or

$$I=1/R \qquad (5)$$

A very important consideration in any economic analysis is what is an adequate return rate. This is often determined by company policy. In the absence of such a policy, a reasonable approach is to compare the risk of the proposed incremental investment with the company average return on net assets and with that of the plant or project, if known, weighing the relative (1) time pattern of earnings, (2) anticipated duration of earnings, and (3) certainty that the earnings will occur.

The per cent return on net assets of leading manufacturing corporations in 1955 varied between 5.5 and 29.1, the average being about 15. Generally speaking, return rates may be expected to be between 0.10 and 0.30 for most new construction in industrial plants.

Earnings that start immediately at a high level, continue for a long period, and are certain to occur without question, are typical of a lower-risk investment. On the other hand, earnings that are estimated to begin some time in the future, may not continue thereafter through the economic life of the investment, and which are less certain to occur, are of the higher-risk type. The lower-risk investment might reasonably have a somewhat lower return rate than the average, whereas, the higher-risk investment might have one a little higher to much higher than average.

To be specific, taking the average return rate of 0.15, the lower-risk investments might be in the 0.10–0.15 range, the medium-risk investment between 0.15–0.20, and the higher-risk investments in the 0.20-0.25 range. These values should be adjusted to the particular case under consideration.

Analysis of Power-System Reliability

The reliability of an industrial power system depends primarily on:

1. The quality of the supply and the supply system.
2. The quality of the system elements.
3. The circuit arrangement of the plant system.
4. The quality of the relay protective system.

Consideration should be given to each of

Table I. Values of a_n Where $a_n=r+\frac{r}{(1+r)^n-1}$

No. of Years, n	Rate of Return, r												
	0.10	0.12	0.14	0.15	0.16	0.18	0.20	0.22	0.24	0.25	0.26	0.28	0.30
1	1.100	1.120	1.140	1.150	1.160	1.180	1.200	1.220	1.240	1.250	1.260	1.280	1.300
2	0.576	0.592	0.6073	0.6151	0.6230	0.6387	0.6545	0.6705	0.6864	0.6944	0.7025	0.7186	0.7348
3	0.402	0.416	0.4307	0.4380	0.4453	0.4599	0.4747	0.4897	0.5047	0.5123	0.5199	0.5352	0.5506
4	0.3155	0.329	0.3432	0.3503	0.3574	0.3717	0.3863	0.4010	0.4159	0.4234	0.4310	0.4462	0.4616
5	0.2638	0.277	0.2913	0.2983	0.3054	0.3198	0.3344	0.3492	0.3643	0.3719	0.3795	0.3949	0.4106
6	0.2296	0.243	0.2572	0.2642	0.2714	0.2859	0.3007	0.3158	0.3311	0.3388	0.3466	0.3624	0.3784
7	0.2054	0.2191	0.2332	0.2404	0.2476	0.2624	0.2774	0.2928	0.3084	0.3163	0.3243	0.3405	0.3569
8	0.1874	0.2013	0.2156	0.2229	0.2302	0.2452	0.2606	0.2763	0.2923	0.3004	0.3086	0.3251	0.3419
9	0.1736	0.1877	0.2022	0.2096	0.2171	0.2324	0.2481	0.2641	0.2805	0.2888	0.2971	0.3141	0.3312
10	0.1628	0.1770	0.1917	0.1993	0.2069	0.2225	0.2385	0.2549	0.2716	0.2801	0.2886	0.3059	0.3235
11	0.1540	0.1684	0.1834	0.1911	0.1989	0.2148	0.2311	0.2478	0.2649	0.2735	0.2822	0.2998	0.3177
12	0.1468	0.1614	0.1767	0.1845	0.1924	0.2086	0.2253	0.2423	0.2597	0.2685	0.2773	0.2953	0.3135
13	0.1408	0.1557	0.1712	0.1791	0.1872	0.2037	0.2206	0.2379	0.2556	0.2645	0.2736	0.2918	0.3102
14	0.1358	0.1509	0.1666	0.1747	0.1829	0.1997	0.2169	0.2345	0.2524	0.2615	0.2707	0.2891	0.3078
15	0.1315	0.1468	0.1628	0.1710	0.1794	0.1964	0.2139	0.2317	0.2499	0.2591	0.2684	0.2871	0.3060
16	0.1278	0.1434	0.1596	0.1680	0.1764	0.1937	0.2114	0.2295	0.2479	0.2572	0.2666	0.2855	0.3046
17	0.1247	0.1405	0.1569	0.1654	0.1740	0.1915	0.2094	0.2278	0.2464	0.2558	0.2652	0.2843	0.3035
18	0.1219	0.1379	0.1546	0.1632	0.1719	0.1896	0.2078	0.2263	0.2451	0.2546	0.2641	0.2833	0.3027
19	0.1196	0.1358	0.1527	0.1613	0.1701	0.1881	0.2065	0.2252	0.2441	0.2537	0.2633	0.2826	0.3021
20	0.1175	0.1339	0.1510	0.1598	0.1687	0.1868	0.2054	0.2242	0.2433	0.2529	0.2626	0.2820	0.3016

Table II. Typical Probable Equipment Failure Rates

Equipment	Probable Annual Failure Rate*
Cable { 2,400-volt and up, per mile	0.133
Cable { 600-volt and below, per mile	0.072
Cable joints	0.005
Circuit breakers	0.010
Transformers	0.003
Transformer with primary switch	0.004
Large motors	0.067

* Wherever available, actual service experience should preferably be used. Actual failure rates will vary widely, depending on quality of equipment, installation, maintenance, and on service conditions.

the foregoing points in determining the reliability requirements.

PROBABILITY

Based on service records of each type of equipment, the probability of failure of each may be found. Table II gives typical probable annual failure rates of equipment. These values may vary widely, depending on the quality of equipment and on installation conditions. Wherever available, actual service records should preferably be used.

In Table III the estimated forced outage times, or the electrical downtimes after equipment failure are given. Much has been written on the use of probability, but little is available in the industrial field.[4-8] By using probability methods, the relative reliability of different circuit or equipment arrangements may be determined.

A set of events is said to be independent if the occurrence of any one of them is not influenced by the occurrence of the others. The probability of failure of one kind of equipment is independent of that of another kind of equipment. Failure rates are generally expressed in terms of failures per year per unit of equipment. Whether or not a piece of equipment is in service has very little effect on the probability of failure. Although failures are not likely to occur when the equipment is out of service, incipient failure may develop during this period. It is assumed here that equipment failures are independent events, although there is some doubt on this point.[4] Using the principles given in Appendix II, the probability of forced outage, defined as the average number of unplanned out-of-service hours per year divided by 8,760, for any combination of elements or circuit arrangement may be found.

Generally speaking, a power outage may cause a loss in three ways. One is proportional to the number of outages, excluding, in many cases, the outages of shorter duration since these may not cause a loss. Those of longer duration may be divided according to duration, each having different values for resulting losses. Such outages may cause damage to plant facilities or may spoil the quality of products being manufactured.

A second kind of power outage loss is proportional to outage duration. It causes lost production. The value of lost production is real, except when the production requirements can be made up after service is restored. The third kind of power outage loss depends on plant startup time after service is restored. This loss is proportional to the number of outages and the plant start-up time.

Thus, the probable gross loss, E, which the plant sustains per year because of electrical failures is

$$E = fL_f + 8{,}760qL_p + fL_pT_s \qquad (6)$$

where

f = probable annual failure rate having outage time long enough to cause fixed loss
L_f = fixed loss per failure, f
L_p = production loss per hour of plant downtime
T_s = time required to start up plant
q = probability of being on forced outage

Take a simple example. If there are two elements of a system in series supplying a load, the failure of either will cause a forced outage. Assume that one is expected to fail once in 10 years and that 48 hours are required to effect repairs and to restore service. Assume that the other is expected to fail twice in 10 years and that 2 hours are required to effect repairs and to restore service. The combination will sustain three failures in 10 years and the service to the load will be interrupted for 52 hours in the 10-year period.

Table III. Estimated Typical Electrical Downtime After Equipment Failure

Equipment	Hours Downtime*
Cable failure, spare length on hand	28
Cable failure, run too long for spare length	41
Cable termination { 2,400-volt and below	10
Cable termination { 6,900-volt	14
Cable termination { 26,400-volt	35+
Circuit breaker	4–8
Transformer	360–720
Motor { 0–100 hp (horsepower) replacement	24
Motor { 0–100 hp repair	170
Motor { 100–500 hp replacement	48
Motor { 100–500 hp repair	240–360
Motor { above 500 hp replacement	72
Motor { above 500 hp repair	400–600

* Actual hours downtime will vary widely. Wherever available actual service records should be used.

If we let

$f_a = 1/10 = 0.10$, probable annual failure rate of first element
$f_b = 2/10 = 0.20$, probable annual failure rate of second element

then,

$$f_{\text{total}} = f_a + f_b \qquad (7)$$

$$= 0.10 + 0.20 = 0.30 \qquad (8)$$

Similarly, let

S_a, S_b = shutdown or forced outage hours per failure of first and second elements respectively
q_a, q_b = probable annual shutdown or forced outage hours per year divided by 8,760, that is, the probability of being on forced outage

Table IV. Assumed System Probable Failure Rates (1960–1969)

Item	System Elements: Description	Quantity	f, Per Unit	Total f
	Case 1—Simple Radial System. See Fig. 3			
a_1	Supply	1	3.000	3.000
a_2	Supply	1	0.100	0.100
Total a				3.100
b_1	Switch	1	0.001	0.001
b_2	Feeder circuit breaker	1	0.010	0.010
b_3	Feeder cable	1/2 mile	0.133	0.067
Total b				0.078
c_1	Transformer and switch	1	0.004	0.004
c_2	Secondary circuit breaker	1	0.010	0.010
Total c				0.014
Total				0.192 or 3.192
	Case 4—Primary-Selective System. See Fig. 6			
b_1	Power-supply switch	1	0.001	0.001
b_2	Switch transfer	1	0.244	0.178
c	Total			0.014
Total				0.259
	Case 6—Secondary-Selective System. See Fig. 8.			
c_3	Tie circuit breaker	1	0.010	0.010
c_4	Breaker transfer	1	0.258	0.192
Total				0.202

Table V. Assumed Outages and Losses per Failure (1960–1969)

Item	Element S, Hour	Plant T_s, Hour	Loss Per Failure, M$, L_f
Case 1—Simple Radial System. See Fig. 3			
a_1	0.5	0	0
a_2	24	20	5
b_1	48	20	5
b_2	8	20	5
b_3	48	20	5
c_1	500	20	5
c_2	8	20	5
Case 4—Primary–Selective System. See Fig. 6			
b_1	48	20	5
b_2	1	0	0
c_1	500	20	5
c_2	8	20	5
Case 6—Secondary–Selective System. See Fig. 8			
c_3	8	20	5
c_4	0.000+	0	0

Table VI. Assumed Plant Hourly Production Loss

Years	Hourly Loss, M$, L_p
1960–1964	1.0
1965–1969	1.3

Table VII. Probable Annual Forced Outage Times and Plant Fixed and Start-Up Losses (1960–1969)

Item	fS (8,760 q)	Yearly Loss M$ Fixed fL_f	Yearly Loss M$ Start-Up fT_sL_p
Case 1—Simple Radial System. See Fig. 3			
a_1	1.50	0	0
a_2	2.40	0.50	2.0
b_1	0.05	0.005	0.02
b_2	0.08	0.05	0.20
b_3	3.36	0.35	1.40
c_1	2.00	0.02	0.08
c_2	0.08	0.05	0.20
Total	9.47	0.975	3.90
Case 4—Primary–Selective System. See Fig. 6			
b_1	0.05	0.005	0.02
b_2	0.18	0	0
c	2.08	0.070	0.28
Total	2.31	0.075	0.30
Case 6—Secondary–Selective System. See Fig. 8			
c_3	0.08	0.4	0.20
c_4	0	0	0
Total	0.08	0.4	0.2

Now,

$$q_a = f_a S_a = 0.10 \times 48/8{,}760 \quad (9)$$

$$q_b = f_b S_b = 0.20 \times 2/8{,}760 \quad (10)$$

Hence,

$$q_{total} = (4.8 + 0.4)/8{,}760 = 5.2/8{,}760 \quad (11)$$

$$= 0.000594 \quad (12)$$

If alternate supplies or circuits are available to a load, the failure rate will be greater because of the added complexity, but the probability of being on forced outage will usually be reduced because of the reduction in time required to restore service. In Appendix II, it is seen that squared q terms, or products of q terms, are generally negligible. In what follows, it is assumed that such terms are negligible.

Typical Case Study

Using the principles just described, the general equations of power system reliability for six simplified but typical systems are developed in Appendix III. For each of the six, reliability is expressed in terms of (1) the probable number of failures per year, and (2) the fraction of the year that the system is likely to be on forced outage, that is, the probability of forced outage.

Example

To illustrate the application of the general equations developed in Appendix III, a simple example is worked out. In Table IV, typical failure rates are assigned to the elements of systems shown in Figs. 3, 6 and 8, and the probable total system failure rates are calculated.

Table V lists (1) assumed outages resulting from equipment failures, both to repair the failure and restore electrical service, and to restart the plant; and (2) the fixed loss per failure. Table VI gives values assumed for plant hourly production loss during forced power system outages plus plant startup time after a failure. It was assumed that plant production losses increase after the first 5 years. This may arise because of an expected increase in unit value of product or in production.

Table VII gives probable annual power system forced outage time because of failures of equipment, and the yearly plant fixed and start-up losses. Table VIII summarizes the total annual losses that result from system outages. These losses are made up of three elements: (1) plant production losses, (2) fixed plant losses, and (3) plant startup losses.

The equivalent uniform annual value, or levelized value of the gross earnings (or losses in these cases), E, may be found by the equation

$$E = a_n \sum_{n=1}^{n=n} E_n / s_n \quad (13)$$

In this equation it is assumed that the investment is made at the beginning of 1959 one year before the plant begins production. The system gross losses, as given in Table VIII, may be levelized as indicated in equation 13, where n, the economic life of the investment is 11 years, the plant being in production for 10 years. However, because the losses are relatively uniform, it is simpler to levelize them as shown in Appendix IV.

Next, in Appendix IV, the different gross earnings of the three power systems are calculated. The investment charge factor, R, of equation 2 is then calculated. It was found to be 0.423 for the values assumed. Lastly, in Appendix IV, the amounts that can be justified to improve power-system reliability in this typical example are calculated. In an actual system study, the number of system elements will usually increase and actual values may be much different from those assumed in the example.

Conclusions

Equation 1 may be used to determine the incremental investment that is justified when it produces gross earnings E. Through the use of probability principles, the probability of power system failures and forced outages may be calculated. General equations for power systems reliability are given in Appendix III. The value of plant losses resulting from forced outages may be calculated using the principles of these general equations, assigning suitable values of equipment failure rates, forced outage time per failure, fixed plant loss per failure, and production loss per hour of forced outage and per hour of plant start-up time.

With this information, equation 1 may be used to determine how much increased

Table VIII. Summary of Annual Plant Losses from Forced Outages

System	Year	Yearly Shutdown Losses, M$ Electrical Failure, 8,760 q = fSL$_p$	Fixed, fL_f	Start-Up fT_sL_p	Total (−E)
Simple radial, Case 1	1960–1964	9.47	0.975	3.90	14.35
	1965–1969	12.30	0.975	5.07	18.35
Primary selective, Case 4	1960–1964	2.31	0.075	0.30	2.69
	1965–1969	3.00	0.075	0.39	3.47
Secondary selective, Case 6	1960–1964	0.08	0.400	0.20	0.68
	1965–1969	0.10	0.400	0.26	0.76

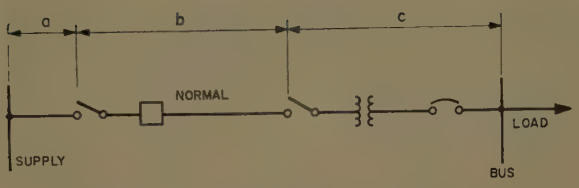


Fig. 3. Case 1—Simple radial system

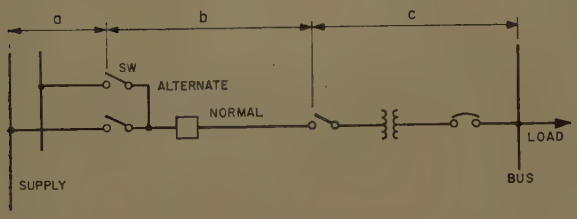


Fig. 4. Case 2—Radial system with selective independent supplies

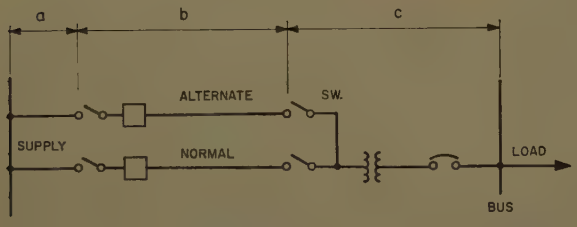


Fig. 5. Case 3—Primary selective system with single supply source

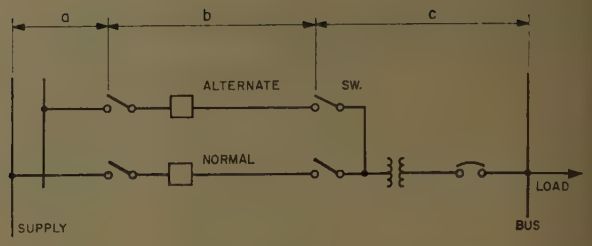


Fig. 6. Case 4—Primary-selective system with independent supplies

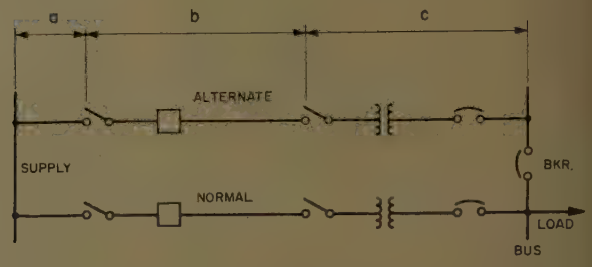


Fig. 7. Case 5—Secondary-selective system with single supply source

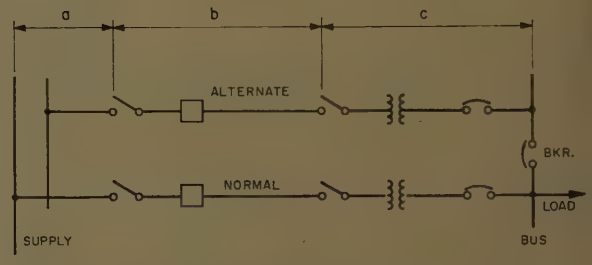


Fig. 8. Case 6—Secondary-selective system with independent supplies

investment is justified to improve power-system reliability. The technique described in this paper enables the question to be answered: Is the extra investment required for a secondary-selective system over a radial system justified, or is the use of a higher quality system element justified?

Appendix I. Derivation of Equation for Justifiable Investment

Annual net income is found by deducting from the gross revenues all operating expenses, depreciation, maintenance, state taxes, insurance, and income taxes. Thus, net income is

$$N=(G-X-eI-j_nI)(1-t) \tag{14}$$

Since cash flow is defined as net income plus depreciation

$$C=N+D \tag{15}$$

$$=(G-X-eI-j_nI)(1-t)+j_nI \tag{16}$$

But C is an annual series of sums which will give a return rate on the investment I plus an amount compounded at rate r for n years, the economic life of the project, sufficient to recover the investment. Hence:

$$a_n=C/I \tag{17}$$

the capital recovery factor. Substituting $C=a_nI$ into equation 16, and solving for I, there results then the following equation:

$$I=E/R \tag{18}$$

where:

$$R=\frac{a_n-j_nt}{1-t}+e \tag{19}$$

$$E=G-X \tag{20}$$

Appendix II. Probability Principles

For a set of, say, two elements in series, the failure of either one of which will take the series out of service, the probability that both will remain in service, or will not fail, is $p=p_1p_2$; where $p_1=(1-q_1)$, $p_2=(1-q_2)$, and q_1q_2 is the probability of forced outage, that is, $q_n=f_nS_n/8{,}760$. In this, f is the probable annual failure rate, S_n is the average number of hours the element, circuit, or plant is out of service because of the failure. Thus, the probability of a forced outage of either of the elements is

$$q=1-p_1p_2, \text{ or } q=1-(1-q_1)(1-q_2)$$

$$=(q_1+q_2)+q_1q_2$$

When q_1 and q_2 are less than 0.01, the error in dropping the term q_1q_2 is less than one per cent. Since this applies to most probabilities of circuit or equipment element forced outage, it is generally sufficiently accurate to combine probabilities of forced outage by adding the forced outage rates, and to neglect the terms involving products.

For two elements that are in parallel, whose forced outage probability are q_1 and q_2 respectively, the probability that both will be out of service simultaneously is q_1q_2. The probability that both will be in service is $p=(1-q_1)(1-q_2)$ or $p=1-(q_1+q_2)+q_1q_2$. The probabilities of all possible operating conditions of two elements is given by $(p_1+q_1)\ (p_2+q_2)=p_1p_2+p_1q_2+p_2q_1+q_1q_2=1$. In this, the probability of being in service with both elements is p_1p_2. The probability of being out of service with both elements is q_1q_2. The probability of being in service with only one element is $p_1q_2+p_2q_1$.

The events, element *1* in service and element *2* out of service, and the event, element *1* out of service with element *2* in service, are known as mutually exclusive events. When events are mutually exclusive, they cannot occur simultaneously and the occurrence of one prevents the other. The probability of the occurrence of either one of the two mutually exclusive events is equal to the sum of the probabilities of the single events.

Appendix III. General Equations for Power System Reliability

In the following refer to Figs. 3–8. These cover six simplified typical power systems. In the following equations, the postsubscripts a, b, c, refer to system elements, and r, p, s, to type of system. Presubscripts refer to the case under study. Symbols used in reference to alternate supplies or circuit elements are primed. Symbols referring to normal supplies or circuit elements are unprimed.

Referring to Fig. 3, a simple radial system (Case 1) is considered. The probability of annual system failures resulting in forced outages is

$$ {}_1f_r = f_a + f_b + f_c \tag{21}$$

The probability of the system being on a forced outage is

$$ {}_1q_r = q_a + q_b + q_c \tag{22}$$

Similarly, for Case 2, a radial system with selective independent supplies, as shown in Fig. 4, the probability of failures and of forced outages is respectively

$$ {}_2f_r = {}_1f_r + f_{sw}' + f_a' q_a \tag{23}$$

$$ {}_2q_r = q_a{}^2 + q_b + q_c + f_{sw}' (S_{sw}'/8{,}760) + f_a (S_{\text{transfer}}/8{,}760) \tag{24}$$

$$ \approx q_b + q_c + q_{sw}' + q_{\text{transfer}} \tag{25}$$

For Case 3, a primary-selective system with a single supply source, as shown in Fig. 5, the corresponding equations are

$$ {}_3f_p = {}_1f_r + f_{sw}' + (f_b')(q_a + q_b) \tag{26}$$

$$ {}_3q_p = q_a + q_b{}^2 + q_c + q_{sw}' + f_b (S_{\text{transfer}}/8{,}760) \tag{27}$$

$$ \approx q_a + q_c + q_{sw}' + q_{\text{transfer}} \tag{28}$$

For Case 4, a primary-selective system with independent supplies, shown in Fig. 6:

$$ {}_4f_p = {}_1f_r + f_{sw}' + (f_a' + f_b')(q_a + q_b) \tag{29}$$

$$ {}_4q_p = (q_a + q_b)^2 + q_c + q_{sw}' + q_{\text{transfer}} \tag{30}$$

$$ \approx q_c + q_{sw}' + q_{\text{transfer}} \tag{31}$$

In Case 5, a secondary-selective system with a single supply, shown in Fig. 7:

$$ {}_5f_s = {}_1f_r + f_{bkr} + (f_b' + f_c')({}_1q_r) \tag{32}$$

$$ {}_5q_s = q_a + (q_b + q_c)^2 + q_{bkr} + (f_b + f_c)(S_{bkr}/8{,}760) \tag{33}$$

$$ \approx q_a + q_{bkr} + q_{\text{transfer}} \tag{34}$$

Finally, in Case 6, a secondary-selective system with independent supplies shown in Fig. 8, the corresponding equations are

$$ {}_6f_s = {}_1f_r + f_{bkr} + f_r'({}_1q_r) \tag{35}$$

$$ {}_6q_s = {}_1q_r{}^2 + q_{bkr} + {}_1f_r (S_{\text{transfer}}/8{,}760) \tag{36}$$

$$ \approx q_{bkr} + q_{\text{transfer}} \tag{37}$$

Appendix IV. Computation of Levelized Annual Losses, Differential System Earnings, Investment Charge Factor and Justifiable Investment

Levelized Annual Losses

For Case 1—Radial system. See Fig. 3 and Table VIII.

$$ -E_1 - = 14.35 + \frac{(18.35 - 14.35)(a_{11})}{a_5 s_6} - a_{11}\frac{14.35}{s_1} \tag{38}$$

$$ = 14.35 + \left(\frac{4.00}{0.334(2.986)} - \frac{14.35}{1.20}\right) \times (0.2311) \tag{39}$$

$$ = \text{M\$ } 12.51 \tag{40}$$

For Case 4—Primary-selective system. See Fig. 6 and Table VIII.

$$ -E_4 = 2.69 + \left(\frac{0.78}{0.334(2.986)} - \frac{2.69}{1.20}\right)(0.2311) \tag{41}$$

$$ = \text{M\$ } 2.35 \tag{42}$$

For Case 6—Secondary–selective system. See Figure 8 and Table VIII.

$$ -E_6 = 0.68 + \left(\frac{0.08}{0.334(2.986)} - \frac{0.68}{1.20}\right)(0.2311) \tag{43}$$

$$ = \text{M\$ } 0.567 \tag{44}$$

Differential System Earnings (Negative Losses)

For Case 4 over Case 1

$$ E_{4-1} = E_4 - E_1 \tag{45}$$

$$ = \text{M\$ } (-2.35 + 12.51) \tag{46}$$

$$ = \text{M\$ } 10.16 \tag{47}$$

Similarly for Case 6 over Case 1

$$ E_{6-1} = \text{M\$ } (-0.567 + 12.51) \tag{48}$$

$$ = \text{M\$ } 11.94 \tag{49}$$

And for Case 6 over Case 4

$$ E_{6-4} = \text{M\$ } (-0.567 + 2.35) \tag{50}$$

$$ = \text{M\$ } 1.78 \tag{51}$$

Calculation of the Investment Charge Factor

The value of R in equation 2 is

$$ R = \frac{r}{1-t} + e + \frac{d_n - j_n t}{1-t} \tag{52}$$

The following values are assumed as typical for the plant investment,

$$ \begin{aligned} r &= 0.20 \\ t &= 0.52 \\ e &= 0.050 \\ n &= 11 \\ j_n &= \frac{1}{n-1} = 0.1 \end{aligned} \tag{53}$$

Substituting these values, there results

$$ d_n = \frac{0.20}{1.20^{11} - 1} \tag{54}$$

$$ = 0.031 \tag{55}$$

$$ R = \frac{0.20}{1-0.52} + 0.50 + \frac{0.031 - 0.1\,(0.52)}{1-0.52} \tag{56}$$

$$ = 0.417 + 0.050 - 0.044 \tag{57}$$

$$ = 0.423 \tag{58}$$

Computation of Justifiable Investment

The incremental investment justified to obtain increased power system reliability is, from equation 1, for Case 4 over Case 1,

$$ I_{4-1} = E_{4-1}/R \tag{59}$$

$$ = \text{M\$ } 10.16/0.423 \tag{60}$$

$$ = \text{M\$ } 24.0 \tag{61}$$

For Case 6 over Case 1,

$$ I_{6-1} = \text{M\$ } 11.94/0.423 \tag{62}$$

$$ = \text{M\$ } 28.2 \tag{63}$$

For Case 6 over Case 4,

$$ I_{6-4} = \text{M\$ } 1.78/0.423 \tag{64}$$

$$ = \text{M\$ } 4.20 \tag{65}$$

References

1. Principles of Engineering Economy (book), E. L. Grant. The Ronald Press, New York, N. Y., 1930, 1938, and 1950.

2. Measuring the Productivity of Capital, Joel Dean. *Harvard Business Review*, Cambridge, Mass., Jan.-Feb. 1954, pp. 120–30.

3. Interest Rate of Return for Capital Expenditure Evaluation, James B. Weaver, Robert J. Reilly. *Chemical Engineering Progress*, New York, N. Y., vol. 52, no. 10, Oct. 1956, pp. 405–12.

4. The Use of Probability in the Design and Operation of Secondary Network Systems, N. M. Neagle, D. R. Nelson. *AIEE Transactions*, vol. 74, pt. III, Aug. 1955, pp. 567–73.

5. Higher Mathematics for Engineers and Physicists (book), I. S. and E. S. Sokolnikoff. McGraw-Hill Book Company, Inc., New York, N. Y., 1934.

6. The Probability of Coincident Primary-Feeder Outages in Secondary Networks, David N. Reps. *AIEE Transactions*, vol. 73, pt. III-B, Dec. 1954, pp. 1467–78.

7. Coincident-Outage Probability in Secondary-Network Vaults, David N. Reps. *Ibid.*, vol. 74, pt. III, Aug. 1955, pp. 580–86.

8. Probability Calculations for System Generation Reserves, Carl Kist, G. J. Thomas. *Electrical Engineering*, vol. 75, Sept. 1956, pp. 806–11.

Discussion

B. E. Crawford (Esso Standard Oil Company, Baton Rouge, La.): Mr. Dickinson's paper is an excellent comprehensive approach to the problem facing the electrical engineer today in deciding to what extent money should be spent to provide additional increments of reliability to a particular load. At Esso's Baton Rouge Refinery similar economic evaluations have been performed for some time. This paper provides briefly most of the considerations necessary to fully evaluate not only the factual data but the proper assumptions and estimates to arrive at a practical solution as to the worth of the electric system

to the particular industry. With the high cost of downtime in many industries, information such as this will show many engineers that a more modern, reliable electric system is often justified.

Although this paper deals primarily with the reliability of the electric system, it gives no consideration to maintenance. The quality of the system is very dependent on maintenance, and the degree of maintenance is usually closely associated with the system design. With large systems in continuous operation, maintenance considerations may be a major concern in system design. Therefore, in determining system design through economic evaluation, consideration must also be given to how maintenance will be performed. For instance, one of the economic benefits derived from a secondary selective-type substation is the ease of performing certain types of maintenance with no loss in production.

The author points out the idea of independent events as associated with probability. The experience at the Baton Rouge Refinery indicates there is much doubt that the type of equipment failures under consideration here are always independent events. However, such dependent failures are usually due to unknown or unpredictable factors, and it is not practicable to consider such failures. We normally design on the basis that we will not have two emergency conditions to occur during the same period of time on the same system.

One of the most difficult items to derive for use in economic evaluation is probable equipment failures. As stated in the paper, actual records of such failures are the best source of information. However, we have found that constant changes to new and better equipment in new construction does not provide long time records suitable for accurately determining this information. We have also felt that this one factor has caused us to put less emphasis on actual failure rate. It would be interesting to know the source of the failure rates given in Table II.

We feel that the method outlined here is the most acceptable method we know of in helping to evaluate investments. In addition to this method, however, we feel past experiences and good judgment definitely should be used in the final analysis as to how much additional money should be expended for improved reliability.

D. S. Brereton (General Electric Company, Schenectady, N. Y.): This paper contains one of the most complete presentations of the economic evaluation of the modern industrial system. One of its principal attributes is that it draws attention to the fact that such an evaluation is possible. The use of the principles of Mr. E. L. Grant have been effectively modified for the power system field.

One of the important concepts that relate to the final analysis is the material given in Table II. It is the user, and not the manufacturer, that has the most complete information on the performance of equipment. It appears that the numbers included in this Table are relatively in the right proportion with the exception of the material dealing with cable joints and circuit breakers. I believe it possible that these two numbers are reversed. It would further be helpful to note, under the subject of transformers, whether liquid-filled or dry-type units are concerned. Although the transformer has the lowest failure rate, if the number given applies to a liquid-filled transformer, the dry-type transformer would probably be in the neighborhood of 0.004.

Mr. Dickinson has done a very capable job of effectively grouping together material on this subject. It is hoped that the AIEE Industrial Power Systems Committee will further document and expand the material given in this paper.

H. B. Backenstoss (Jackson & Moreland, Inc., Boston, Mass.): The author has undertaken a commendable and difficult task in isolating and defining in dollar terms the specific item of reliability in an industrial electric system. He has made an important contribution on an important subject.

Although the application of probability to electric system design and the evaluation of outage losses are the aim of this paper, the section dealing with the technique for economic evaluation has the prominence of first appearance. The development of the formulas for this technique raises some questions which bear examination even at the expense of neglecting other aspects of this worthwhile paper. The following discussion is confined entirely to the evaluation technique.

The discounted-cash-flow method used by the author and described in several of his references is a mathematically rigorous method for determining the interest rate of return. Basically it is a statement of the identity underlying all money lending calculations, namely, that the present worth of all future interest payments plus repayments of principle must exactly equal the initial value of the loan. In the discounted-cash-flow method of analyzing an investment, the earnings on the investment (interest) and annual depreciation accruals (repayments of principal) are lumped together as total cash flow and not individually identified. The interest rate for the present worth arithmetic which satisfies the identity of these payments with the original investment is the rate of return earned by that investment.

A project's cash flow has the advantage of being a definite measurable amount. It very usefully by-passes the problem of depreciation which has caused endless confusion as to which should be used, i.e., tax-allowable, straight line, or sinking fund. It can be correctly determined by two different computations, either by steps typical of those used by a company in computing its Federal Income tax obligation, or by steps typical of those it uses in preparing its income statement for stockholders. Due to the possibility open to a company to use accelerated depreciation in determining its taxable income and normal depreciation in determining its reported income, the two methods are not identical, although they both show exactly the same cash flow, i.e., net income plus depreciation.

The author has used the "tax form" approach in developing expressions for R in the text and in Appendix I. This conceals the tax credit which the difference between tax allowable depreciation and economic depreciation (sinking fund) provides. The importance of this difference is increased where a company uses SYD (sum-of-years digits) or other similar accelerated methods for tax purposes.

The expressions for R may readily be transposed to a form which reveals the tax credit, as follows:

$$R=\frac{a_n-j_nt}{(1-t)}+e \tag{66}$$

$$=\frac{a_nt}{(1-t)}+\frac{a_n(1-t)}{(1-t)}-\frac{j_nt}{(1-t)}+e \tag{67}$$

$$=a_n+\frac{t}{(1-t)}(a_n-j_n)+e \tag{68}$$

$$=r+d_n+e+\frac{t}{(1-t)}[r-(j_n-d_n)] \tag{69}$$

This expression separates return r from the other components of total carrying charge on investment and also identifies income tax as a separate item in terms of the tax rate t, the return r and a credit related to the difference between tax-allowable depreciation and economic depreciation. Expense e and depreciation d_n are the same as in the author's equation 2. These expressions would all be somewhat different for a company which had substantial debt financing.

It may appear on the surface that, if this formula gives the same result for R as the author's, the distinction between the two is of no moment. In fact, however, there are subtle overtones of some importance.

It is desirable that an economic evaluation, which aims at measuring the earning power of a proposed new investment, adopt the methods which the company itself uses in measuring its own earning power. To use other methods in evaluating a project which, after acceptance, becomes an increment addition to the company's property is apt to be confusing. This is particularly true in the case of management, which is almost surely predisposed to think in terms of the company's own methods when reviewing proposed capital expenditures.

The methods used by a company to measure its earning power are the methods used in preparing an annual income statement, for those methods, although not standard for all companies and subject to discretionary variation within a given company, are the ones used to determine net income. And net income as it appears on a company's income statement is generally regarded as the measure of a company's profitability. Net income as it appears on a tax statement is commonly a much lesser amount. It fails to include the depreciation tax credit exemplified in the last term of equation 4 above. The author's equation 14 defines net income as it appears on a tax statement. In reality, it defines "taxable income less taxes," a quite different figure from net income.

By taking the "income statement" approach in evaluating proposed capital investments one avoids falling into the error of thinking that income taxes should be 52% (the current rate) of income before taxes. The 1956 annual report of Standard

Oil Company of New Jersey furnishes an example:

Net income before taxes	$1,238,500,000	100.0%
Provision for income taxes	430,000,000	34.7%
Net income	$ 808,500,000	65.3%

From equation 14 one might expect income taxes to be 52% of net income before taxes, rather than only 34.7%. Accelerated tax-allowable depreciation is almost surely the primary cause of the great reduction in the tax that is evident here.

One also avoids another pitfall, a second facet of the same problem. It is the pitfall of interpreting the ratio of taxable income less taxes divided by investment as the rate of return r. Referring again to Standard Oil Company's income statement:

Provision for income taxes	$430,000,000	52%
Corresponding taxable income (computed as $430,000,000/0.52)	827,000,000	100%
"Net income" (Actually taxable income less taxes)	$397,000,000	48%
Actual net income	$808,500,000	

The tax credit provided by accelerated depreciation adds directly to net income and is a significant amount. It would be quite incorrect to use the much smaller item of taxable income less taxes as a gauge of profitability.

By separating cash flow a_n into its components, return r and depreciation d_n, in equation 69, one can also see that a plan of low cash flow, but low depreciation due to long life, may provide a higher return and hence be more attractive than a plan of higher cash flow having a high depreciation. This clarifies the author's statement about cash flow. It is not the flow itself but the interest rate of return it provides which determines the economic choice of plans, and that is what he means.

In this connection, it is well to bear in mind that the rates of return between 5.5 and 29.1% reported by the author are not representative values to use for new investments. They are too high, having been inflated by past price rise. They are the earnings rates at current price levels on investments committed at lower price levels. The amount of increase in earnings rate caused by this effect may be difficult to determine. It would appear, however, that 15%, the author's suggested average rate of return, is appreciably higher than the rate at which past investments were made and were believed attractive.

Another factor inherent in the development of the formulas in this paper will probably be evident to most of those who make economic comparisons. By treating only the differential investment and differential annual costs, the formulas presuppose that the carrying charge rate for the increment is the same as the carrying charge rate for the base investments. Although this is probably the case in most electric design alternatives it need not invariably be so. One plan may have a longer life or different degree of risk and its total carrying charge rate will differ from that of a second plan. In this event the two plans cannot accurately be compared by the formulas developed here. Where the carrying charge rate varies for any reason it becomes necessary to calculate annual charges for each plan on either a year-by-year or a levelized basis and obtain differences in cost by direct subtraction.

The author is to be congratulated for his interest and his efforts in placing the economic evaluation of reliability in an electric system on a sound mathematical basis, instead of a basis of personal preferences and guesswork which have often heretofore been the principal guiding influences.

W. H. Dickinson: The author wishes to express his appreciation to Mr. Crawford, Mr. Brereton, and Mr. Backenstoss for their discussions. The point raised by Mr. Crawford regarding the importance of maintenance is a good one. Contrary to his impression, the equation $I=E/R$, developed in the paper, does take maintenance into account. In addition to normal maintenance, even losses resulting from plant downtime which are required to make additions to the electric system, may be accounted for. The value of E, the gross earnings, equals gross revenues less variable expenses X. Thus X covers variable maintenance expenses as well as other variable expenses. The investment charge factor R includes the expense rate e, which covers fixed expenses, including fixed maintenance. Of course, if the values of variable maintenance are not uniform during the life of the project, they must be levelized as described in the paper. Essentially, this consists of finding the uniform annual amounts over the life of the project having a present worth equal to the present worths of the variable maintenance costs.

Mr. Crawford also questioned whether the probability of forced outages resulting from system component failure should be regarded as independent events. After wrestling with this problem for some time, the author concluded that it was more appropriate to treat forced outages as independent events. This is the method used recently[1] in calculating forced outage rates on unit-type installation of boiler and turbine generators. In a discussion of their paper (see reference 4 of the paper) Mr. Neagle and Mr. Nelson stated that they now regarded these events as mutually exclusive, whereas in a previous paper they had treated them being independent events. The author believes that forced outages are not truly mutually exclusive in that the occurrence of one forced outage does not prevent the occurrence of a simultaneous failure.

From a practical point of view, the forced outages rates of electric equipment in industrial power systems are usually low. The forced outage rates of most simple radial systems will usually be less than one per cent. Hence, it will make no significant difference in final results whether forced outage rates or failure rates are treated mathematically as independent events, or as mutually exclusive events. For example, let forced outage rates of one per cent be assigned to two items connected in series. If the probability of failures are regarded as mutually exclusive events, the combined forced outage rate will be found accurately by addition of the forced outage rates, that is, $0.01+0.01=0.02$, or 2%. If, instead, the events are regarded as independent, the probability of being in service is found by multiplying in service probabilities, that is, $(1-0.01)(1-0.01)=1-2(0.01)+(0.01)(0.01)=1=0.02+0.001$. The probability of being on forced outage is found by subtracting this term from one, giving the result $0.02-0.0001$. Dividing the difference, 0.0001 by 0.02 shows that there is only 0.5% difference in the two results. This is well within the accuracy of the forced outage rates themselves.

As Mr. Crawford points out, economic evaluation is not the only consideration in determining power system reliability requirements. There are always intangibles which cannot be evaluated. But past experience should not be one of them, because it should be possible to measure past experience, that is, system failures and their effects on plant earnings, in dollars and cents. If this is done, preconceived notions of what can be justified will be avoided.

Mr. Brereton and Mr. Crawford have both commented on Table II, "Typical Probable Equipment Failure Rates." The sources of these data are (1) published reports of the Edison Electric Institute and given in reference 4 of the paper, (2) AEIC (Association of Edison Illuminating Companies) Report of Committee on Power Distribution 1954–1955, covering cable operation for the period 1943–1953, and (3) scattered and incomplete information from various other sources. None of these sources included information on circuit breaker failure rates. These were estimated. However, as a check, the Edison Electric Institute gives a failure rate of network protectors to close as 0.0292. Failure of a network protector to close is considered more likely than the failure of a circuit breaker. Hence, a circuit breaker failure rate of 0.01 was selected. As to cable joints, failure rates of 0.005 and 0.00159 were reported by the EEI and the AEIC reports respectively. Similarly, failure rates for cable per mile per year are given as 0.068 and 0.072 by the EEI and AEIC reports respectively. In each case, the reported figures were arbitrarily increased to those quoted in Table II, to allow for industrial conditions, which were thought to be more severe than the utility experience on which the EEI and AEIC reports were based. Transformer failure rates were based on the EEI report in which the figure 0.00274 was quoted. These are undoubtedly liquid filled transformers. In Table II this figure was rounded out to 0.003, for a transformer only and to 0.004 for a transformer with a switch. The author has no information on dry-type transformers, but Mr. Brereton's suggested failure rate of 0.004 appears reasonable.

It is unfortunate that there is so little information available from industrial sources. With some digging no doubt some information could be found in the records of maintenance departments. Over-all forced outages of plants should also be available from the record of plant operating personnel, which could be used to check results. The author concurs wholeheartedly with Mr. Brereton's suggestion that the AIEE Industrial Power Systems Committee further document and expand the material given on the paper. A practical way of doing this might be the preparation

of a questionnaire similar to the form prepared on forced outage rates on steam turbines and boilers.[1]

Mr. Backenstoss has confined his discussion to the economic aspects of the paper. He agrees that the discounted-cash-flow method is a mathematically rigorous method for determining the interest rate of return.

Regarding depreciation, far from bypassing, as Mr. Backenstoss indicates, the problem of depreciation, the paper attempts to meet this problem head-on. It provides for the use of sinking-fund depreciation for the economic analysis after computing income taxes, using for tax purposes the type of depreciation actually used by the company in question, and permitted by the income tax authorities. It is sometimes the case that a company uses still a third type of depreciation for financial accounting purposes. For example, it may use accelerated depreciation, such as sum-of-years' digits or declining balance, for income tax purposes, and normal or straight line depreciation for financial accounting purposes. In such cases, a tax equalization reserve or reserve for deferred income taxes is often set up. This third type of depreciation is of no consequence in an economic analysis. The term

$$N=(G-X-eI-j_nI)(1-t) \tag{70}$$

as given in equation 14 of the paper precisely defines the meaning of N. If the term, j_nI, the depreciation used in computing income taxes is also used for financial accounting purposes, and if no capital for the incremental investment I was derived from borrowing, then, and only then, may the term N be properly called "net income after taxes." In the paper this was assumed to be the case for the incremental investment and hence N was called net income. However, this is a loose use of the term, which, as used by accountants, means the amount earned or available for the stockholders after allowing for income taxes, interest payments to lenders, and book depreciation. Book depreciation may differ from that used in computing income taxes. There is no accepted term to describe N. Perhaps "taxable income less income tax (no borrowed funds)" would be a better way to refer to N. Similarly, equation 15 in the paper precisely defines cash flow. It is

$$C=N+D \tag{71}$$

where N is as defined above and D is the j_nI, the income tax depreciation. Neither the question of semantics in referring to N, nor the treatment of the incremental investment on the books of account should be allowed to confuse the issue when making an economic analysis. For this, find the annual cash flows C over the life of project, and from these determine r the rate of return on the investment, as described in the paper.

Mr. Backenstoss states that cash flow, as determined by equation 2, will be the same as net income after taxes plus book depreciation where book depreciation differs from income tax depreciation. This is not correct. For example, assume that the installed cost of a device is \$300, has a 3-year life, and has gross earnings (before depreciation or income tax) of \$200. The first year the straight line depreciation is \$100 and the declining balance depreciation is \$200. If the income tax rate is 50%, the cash flow in the former case is

$$C=(200-100)(0.50)+100 \tag{72}$$

$$=\$150 \tag{73}$$

and in the latter case is

$$C'=(200-200)(0.50)+200 \tag{74}$$

$$=\$200 \tag{75}$$

Mr. Backenstoss has developed in his equation 69 another form for the investment charge factor than those shown in the paper where the two different forms were used:

$$R=\frac{a_n-j_nt}{1-t}+e \tag{76}$$

$$R=\frac{r}{1-t}+e+\frac{d_n-j_nt}{1-t} \tag{77}$$

The form developed by Mr. Backenstoss is as follows:

$$R=r+d_n+e+\frac{t}{1-t}\ [r-(j_n-d_n)] \tag{78}$$

All three of these are algebraic identities. It would appear to be a matter of personal preference which one is used. The author prefers the first form because it is the most concise and in the best form for ease in computation. It is also meaningful when the significance of the terms is understood.

All three of these forms give exactly the same results and therefore give the same measure of the earning power or profitability. The methods described in the paper are intended for use in determining the incremental investment that can be justified. They are not intended for measuring the profitability of over-all company operations. In any event, it is not net income after income taxes but the annual cash flows, as defined in equation 2, or equation 16 in the paper, over the life of the project which should be used in an economic analysis to determine the rate of return. This is at the very heart of the discounted-cash-flow method which Mr. Backenstoss agrees is a mathematically rigorous method.

The paper assumes that investment in the project is not from borrowed funds, that is, that no interest payments are involved. This simplifies the economic analysis and avoids the necessity of determining the percentage of borrowed capital and the interest rates being paid thereon. Because of the tax credit on interest payments, the rate of return is higher when a part of the funds are borrowed. A correction[2] can be made to the rate of return, if a fraction of the investment is actually derived from borrowing. In an industrial company, it is usually the case that the fraction of the investment derived from borrowing is lower, and the rate of return on investment is higher, than that which prevails in public-regulated utilities. As a consequence, the magnitude of the correction for borrowing is usually smaller for an industrial than it is for a utility. It may often be neglected without affecting results significantly.

Mr. Backenstoss cites advantages in taking the income-statement approach rather than the tax-form approach: (1) it gives a better measure of a company's profitability, and (2) it avoids thinking that the income tax rate should be 52%. It may be true that over-all company operations may be more truly shown by an income-statement approach which differs from a tax-form approach. Nevertheless, for an incremental economic analysis, the tax-form approach gives correct results, because it correctly measures the cash flows, as defined by equation 16 in the paper, from which the rate of return on the investment is derived.

Mr. Backenstoss infers that the paper suggests the use of 52% as the income-tax rate. He believes that this is an error since the actual rate is much lower because of accelerated tax-allowable depreciation.

The author believes that the correct tax rate to use in an economic analysis is not the effective tax rate but the maximum rate applicable on incremental pretax earnings. According to present income tax rules, this rate is 52% for taxable earnings in excess of \$25,000. For the larger industrial companies, the maximum rate of 52% would usually be the one used. Most writers on problems involving economic analysis seem to agree with this view because they use 50 or 52% as a tax rate in the examples they cite. Furthermore, in Barrons, August 5, 1957, figures are given for U.S. corporate profits after tax and before tax during the past year. These have a ratio of 0.49, or an indicated average tax rate of 51%.

The figure of 15% for an average rate of return is questioned by Mr. Backenstoss. The rate which was suggested in the paper for economic comparisons was not this, but a rate determined by company policy, or by finding the company average return on their net assets, and adjusting this to the risk of the proposed investment as compared with the average risk. It may also be adjusted to allow for increases in prices, as Mr. Backenstoss feels is required, although the author believes, in view of very substantial expansion programs carried out over the past few years by many industrial companies, at prices close to today's, that this factor is of less importance than Mr. Backenstoss seems to indicate.

As Mr. Backenstoss points out, the comparison of two alternatives using differential earnings presupposes that both alternatives have the same investment charge factor. However, the equation $I=E/R$, as developed in the paper, can be used to compare two alternatives having different investment charge factors. This is done by comparing each with a third alternative of no investment. After computing gross earnings E for each plan, the justifiable investments for each may be found and compared with each other and with the base case of making no investment.

References

1. Forced Outage Rates of High-Pressure Steam Turbines and Boilers, AIEE Joint Subcommittee Report. *AIEE Transactions*, vol. 76, pt. III, June 1957, pp. 338–43.

2. Some New Mathematical Aspects of Fixed Charges, Constantine W. Bary, W. T. Brown. *Ibid.*, pp. 230–49.

Application of a Mechanical Rectifier to a Cyclotron

JACK WARREN
ASSOCIATE MEMBER AIEE

THE first mechanical rectifier to be used in the field of "atom-smashing" has been installed and has completed 8 months of operation. Anticipated high reliability and rectification efficiency has been realized. To date no major operating or maintenance problem has arisen.

In February 1956, a 3,000-ampere 450-volt d-c mechanical rectifier manufactured by the I-T-E Circuit Breaker Company was placed in service at Columbia University's 385 million electron-volt synchrocyclotron at Irvington, N. Y., as the power supply for the cyclotron electromagnet.

Reasons for Using Mechanical Rectifier

In 1953, although the cyclotron had been in operation since 1950, it was decided to obtain a new magnet power supply because of the poor performance record of the motor-generator set then in use, and because cyclotron revisions were being considered which would eventually require a power supply of higher rating.

The motor-generator set consisted of a 900-horsepower 2,300-volt 3-phase synchronous motor, and a 600-kw 300-volt shunt generator. During its 3 years of use, five major breakdowns caused a total loss of $13\frac{1}{2}$ months of cyclotron operating time.

Various manufacturers were invited to suggest a new power supply. Proposals were received for motor-generator sets, ignitrons, a selenium rectifier, and a mechanical rectifier. These devices were compared on the basis of initial cost efficiency, space requirements, cooling requirements, and reliability. In regard to the last requirement, a reliable device meant one that would break down only infrequently and when it did, could be repaired either by the manufacturer or by the cyclotron operating staff in a reasonably short time. Investigation indicated that any of these devices, if well designed and constructed, should be reliable. The reliability of the mechanical rectifier was particularly investigated since this was the newest device. Every user of an I-T-E mechanical rectifier was questioned, and the findings were favorable. In comparing the efficiency of the various devices, consideration was given to fractional load efficiency, since at least the first few years of operation would be at about 40% of the power supply's rated load. On the basis of all the aforementioned considerations, the decision to purchase a mechanical rectifier was made.

Description of Cyclotron Magnet

The steel yoke of the magnet weighs 2,000 tons. It is 33 feet long and 21 feet high. The upper and lower coils for the magnet are each constructed of six spirally wound pancakes of copper bus bar. Each pancake is made up of 5,700 feet of 4-inch-wide by 9/32-inch-thick bus bar. The total copper weight of both coils is 150 tons. Each coil is in a 22-foot-outside-diameter coil tank which resembles a doughnut with a square cross section. These tanks are primarily liquid containers for the cooling oil.

The resistance of the magnet varies from 0.120 to 0.147 ohm, depending upon its temperature. The inductance varies from about 2 henrys to 30 henrys, depending upon the current. The total energy stored in the magnet at 3,000 amperes ($\int_0^{3,000} Lidi$) is about 20,000,000 joules. This high inductance and stored energy were factors which weighed heavily in the design of the rectifier. If, while the magnet were carrying current, its input should open for any reason, dangerous overvoltages would be generated by the magnet. Damages to the magnet could be so severe that repair time, and hence cyclotron shutdown time, could be of the order of a year. Repair costs in case of severe damage could amount to the order of several hundred thousand dollars.

Description of Rectifier Equipment

Fig. 1 shows the schematic diagram for the basic rectifier circuit.

Outdoor Substation

The outdoor substation of the laboratory was expanded to accommodate the a-c switchgear and step-voltage regulator; see Fig. 2. The a-c switchgear unit is of weatherproof construction and contains the main a-c circuit breaker, rated at 1,200 amperes, 3 phase, 15,000 volts, with an interrupting capacity of 500 megavolt-amperes.

Other major components of the outdoor switchgear unit are an auxiliary power transformer, a battery rack and charger and potential and current transformers.

The battery consists of a bank of nickel-cadmium alkaline batteries which supplies 48 volts d-c for tripping power for the main a-c and d-c circuit breakers. The batteries were chosen for a reliable source of tripping power in case of a failure of the one 13.8-kv feeder. An automatic charger keeps these batteries charged. The potential and current transformers are for meters, instruments, and protective relay purposes.

The step-voltage regulator is rated at 239 kva, 13,800 volts, 3 phase, 60 cycles, 16% raise and lower. It is essentially a 3-phase regulating autotransformer of the oil-immersed self-cooling type. It is provided with taps in a regulating (series) winding and a multiposition, electrically operated tap-selector switch which will give automatically a range of voltages in the series winding. Thus, voltage regulation from 16% above to 16% below line voltage is obtained in 32 1% steps.

It should be noted that for most applications of the mechanical rectifier, a step-voltage regulator would not be necessary. The voltage regulation available from the taps on a normal rectifier transformer (about 10%) plus the voltage regulation available from the normal commutating reactors (approximately 50%) is usually ample enough. The reason that a step-voltage regulator was used in this application is that it was desired for a particular rectifier transformer tap to energize the cyclotron magnet at a minimum direct voltage to keep the rate of rise of magnet flux at a minimum. This would minimize the eddy currents developed in the copper dee structures, which in turn would limit undesirable physical stresses due to reaction between the flux set up by the eddy currents and the main flux of the magnet.

Cyclotron Building

Fig. 3 shows the layout of the rectifier equipment in the cyclotron building.

The output of the step-voltage regulator in the outdoor substation is fed into the

Paper 57-700, recommended by the AIEE Industrial Power Rectifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted September 17, 1956; made available for printing May 15, 1957.

JACK WARREN is with Singmaster and Breyer, New York, N. Y., and the Columbia University Nevis Cyclotron Laboratories, Irvington, N. Y.

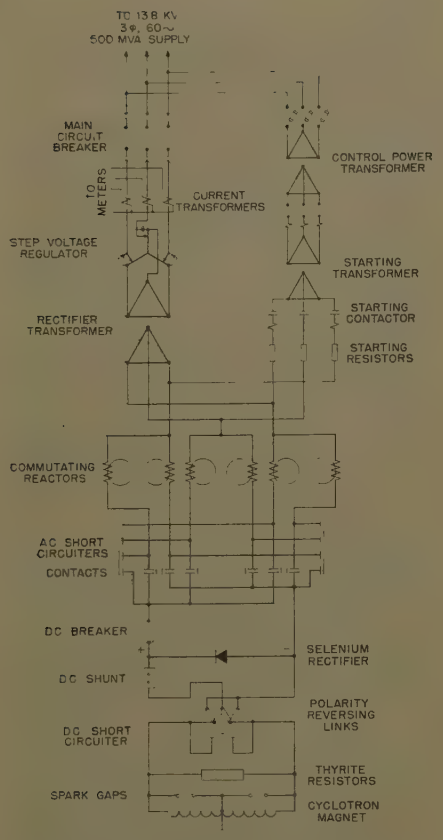


Fig. 1. Basic rectifier circuit, schematic diagram

cyclotron building to the rectifier transformer. The transformer is rated at 1700 kva, 15,900 volts, 3 phase, 60 cycles, 55-degree-centigrade temperature rise at full load, 3.4% impedance. It is Askarel-filled and self-cooled. There are five no-load taps which permits a total reduction of about 11% from maximum voltage. The no-load tap changer is controlled by an operating handle on the transformer tank. It is key interlocked to prevent its operation during load conditions.

The six commutating reactors are mounted in the same tank with the transformer winding. There are three control windings on these reactors in addition to the main load current winding: the break pre-excitation, the flux reversal, and the overlap.

The break pre-excitation winding supplies the reactor core with the current necessary to reverse its flux during the break step, thereby relieving the load winding from this duty and enabling the contact to open while it is carrying close to zero current. This winding, in conjunction with an arc suppressor, an *R-C* circuit across the contact, completes the flux change in the reactor after the contact opens so that the contact has time to separate sufficiently before it sees the inverse voltage.

The flux reversal winding enables the commutating reactor to be used for regulating the rectifier output voltage. This winding, together with the control circuit that feeds it, permits the flux of the reactor core to reverse a controllable amount before the closing of the contact, thereby enabling the time for completion of flux reversal after the closing of the contact, the make step, to be shortened a predictable amount of time. The controlling of the amount of flux reversal prior to a contact closing, and therefore the length of the make step after the contact closes, is similar to the use of phase-shift control of the firing point of grid-controlled mercury-arc rectifiers and ignitron rectifiers. The longer the make step, or phase shift delay, the less will be the output voltage.

The flux reversal windings, besides enabling the commutating reactors to regulate voltage over a wide range, enable the commutating reactors to control cyclotron current to a fine degree. For the sake of cyclotron beam stability, specifications called for the magnet power supply to regulate magnet current automatically to at least $\pm 0.25\%$ of a preset value.

The major components of the automatic current controller are shown in the block diagram in Fig. 4.

The cyclotron magnet current is sensed by a 3,000-ampere 30-millivolt shunt; it has a limit of error of 0.04%. The shunt output is fed to a self-balancing-type potentiometer current recorder-controller with a range span of 600 amperes. Four suppressed zero ranges, 1,000, 1,500, 2,000, and 2,500 amperes, are provided so that current on any one range can be read to about 1 ampere. The accuracy of the instrument is $0.1\% \times$ suppression ratio $+0.3\%$ of range span. The dead band of the instrument is 0.15% of the range span, or 0.9 ampere.

A second slidewire, in addition to that of the self-balancing potentiometer, is provided for control purposes. If the magnet current, as indicated by a black pointer on the instrument scale, deviates from a preset value as selected by a red pointer on the same scale, the sliding contact of the control slidewire moves off center an amount proportional to the deviation.

The position of the control slidewire is converted to current from zero to 5 milliamperes, by an electronic control unit. Proportional band, reset, and rate circuits are provided for adjusting for optimum automatic magnet current control.

The load in which this control unit develops the current is the signal winding of a magnetic amplifier. The d-c output of this magnetic amplifier, the flux reversal current, is fed to another control circuit which feeds the flux reversal winding of the commutating reactor. A decrease of magnet current causes the automatic current regulator system to increase the amount of flux reversal in the commutating reactor core before the contact closes, thereby decreasing the length of the make step, and thereby increasing the rectifier output voltage and hence the cyclotron magnet current.

The third control winding of the commutator reactor is the overlap winding. Overlap time is the time that the contact about to give up the load current and the contact about to take it are both closed. An inherent characteristic of this type of rectifier is that to keep contacts breaking in the middle of their break step, the overlap time must vary with input voltage, phase control, and load. Automatic control of the overlap time, while maintaining the same closing time for contacts, is accomplished by the overlap regulator and circuit. Position of the beak, during the break step, is fed to the overlap regulator control circuit via the overlap winding on the commutating reactor.

High-speed a-c and d-c short-circuiting breakers are provided to protect contacts in case of backfire. These are latched against spring tension in the open position by the flux of a permanent magnet. Tripping is caused by shifting flux by energizing a coil wound on the magnet. This coil is energized by impulse transformers when these sense unbalances between the a-c and d-c systems of the rectifier. The breakers close within about a millisecond of the initiation of the trigger pulse. The d-c short circuiter is also one of several devices for insuring that the cyclotron magnet has a discharge path in case of trouble.

A single-pole 3,000-ampere d-c circuit breaker is in one of the output busses of the rectifier. It is apparent that this breaker cannot by itself open the highly inductive load, as the maximum stored energy of the magnet is approximately 20,000,000 joules and the maximum arc energy that could be dissipated by any d-c breaker is under 1,000,000 joules. And as previously stated interrupting the magnet circuit while it was carrying current would cause the magnet to generate dangerous overvoltages.

Use of the d-c breaker was made possible by a bank of selenium rectifiers located in the breaker cubicle. The bank is connected in reverse polarity as shown in Fig. 1, so that it does not conduct under normal operation. When the d-c breaker is opened, the voltage at the terminals of the load reverses and the energy stored in

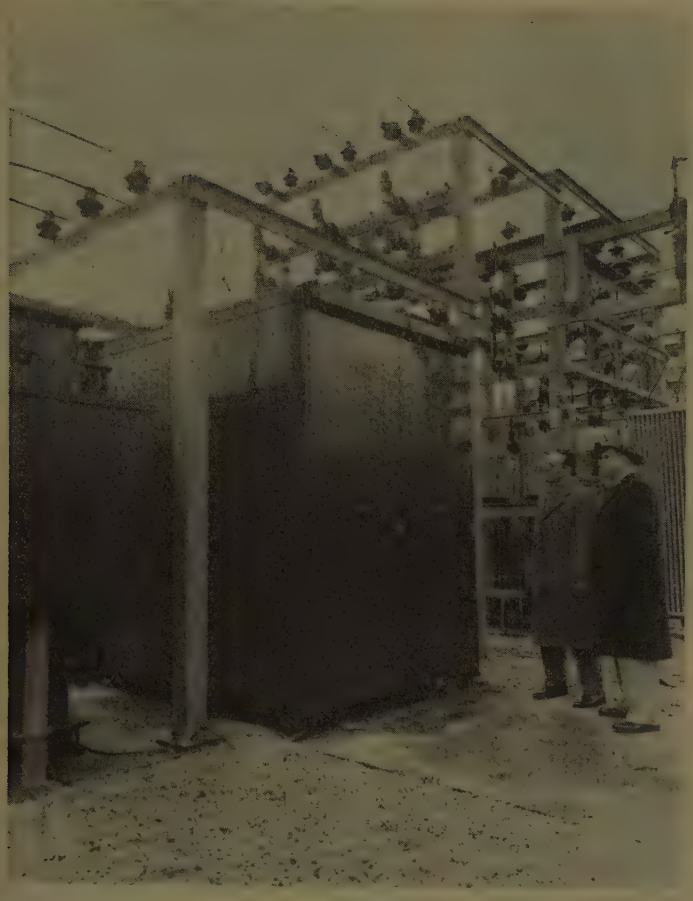


Fig. 2. Substation showing step-voltage regulator at left, a-c switchgear in center

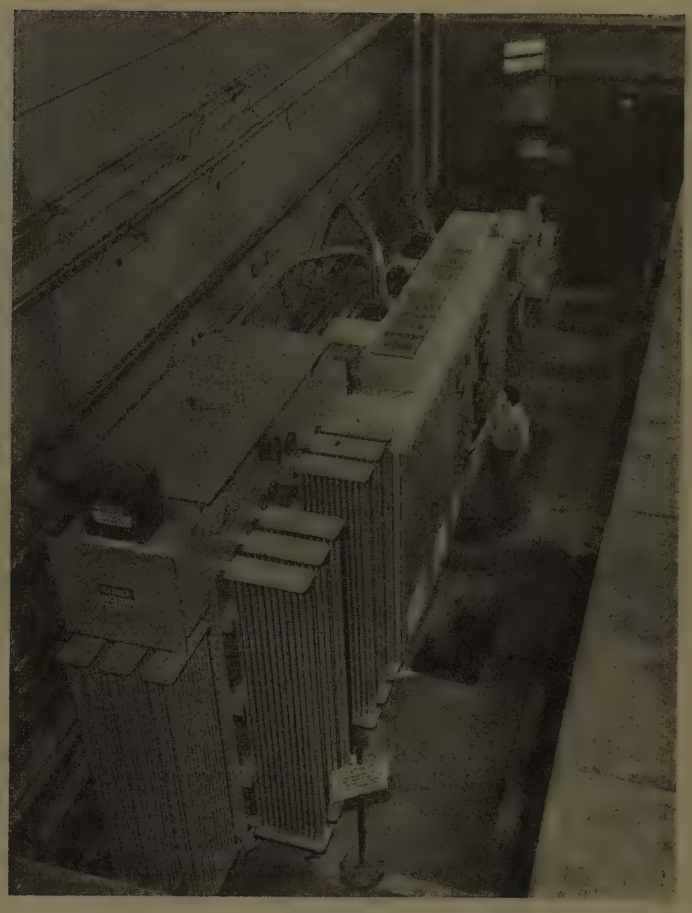


Fig. 3. Mechanical rectifier equipment in cyclotron building

the magnet dissipates itself in the resistances of the magnet and selenium rectifiers. The normal leakage power dissipation at 450 volts will be about 300 watts or 0.23% of rated.

To protect the expensive cyclotron magnet further against any unforeseen contingency, a bank of Thyrite resistors is connected across the copper bus to the magnet. The Thyrite is rated to dissipate the maximum energy of 20,000,000 joules while limiting the maximum induced voltage across the magnet terminals to 1,700 volts. Normal dissipation in the Thyrite at the 450-volt level will be about 3 kw, or 0.23% of rated rectifier load. At the present operating level of 260 volts, the dissipation is only about 100 watts, or 0.02% of the rectifier output.

To protect the magnet still further, a pair of spark gaps is installed at each of the magnet coil terminals to ground. A spark gap consists of two $4^3/_4$-inch-diameter copper disks, 1/2-inch thick, separated by 0.015 inch of mica, all enclosed in a micarta enclosure. Tests have shown that these gaps break down at about 1,700 volts.

It is necessary, at times, to reverse the polarity of the current through the magnet. This is done by first shutting down the rectifier and permitting the magnet current to decay to a low level. Then, when the residual current is low, a special double-pole double-throw "make-before-break" polarity-reversing switch is operated. This switch is located in the d-c breaker cubicle. The door of the cubicle is key interlocked so that the switch cannot be reached unless the d-c breaker is open. A meter at the switch indicates the amount of residual magnet current.

After the switch is operated, the selenium presents inverse resistance to the residual current. Depending upon the value of this inverse resistance and the value of the residual current, a high voltage could be induced by the magnet. To eliminate this possibility, a 300-ohm resistor is permanently connected across the selenium.

Controls

Startup is initated by the starting control switch. This energizes the rectifier mechanism through the starting transformer and starting resistors. The main a-c breaker control switch is then operated which energizes the rectifier through the main transformer. The next step is to energize the magnet by operating the d-c breaker control switch and closing the breaker. This breaker is interlocked so that it cannot be closed unless both the step-voltage regulator and phase control

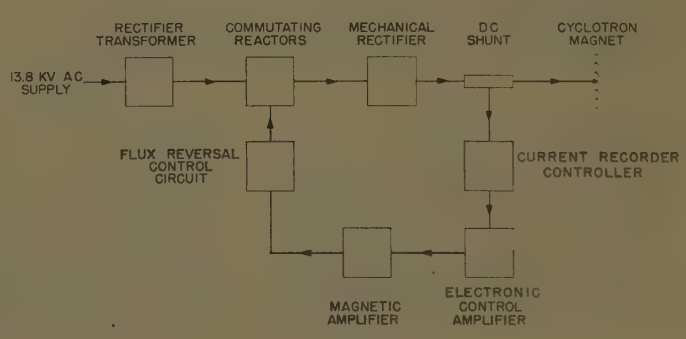


Fig. 4. Automatic current control, block diagram

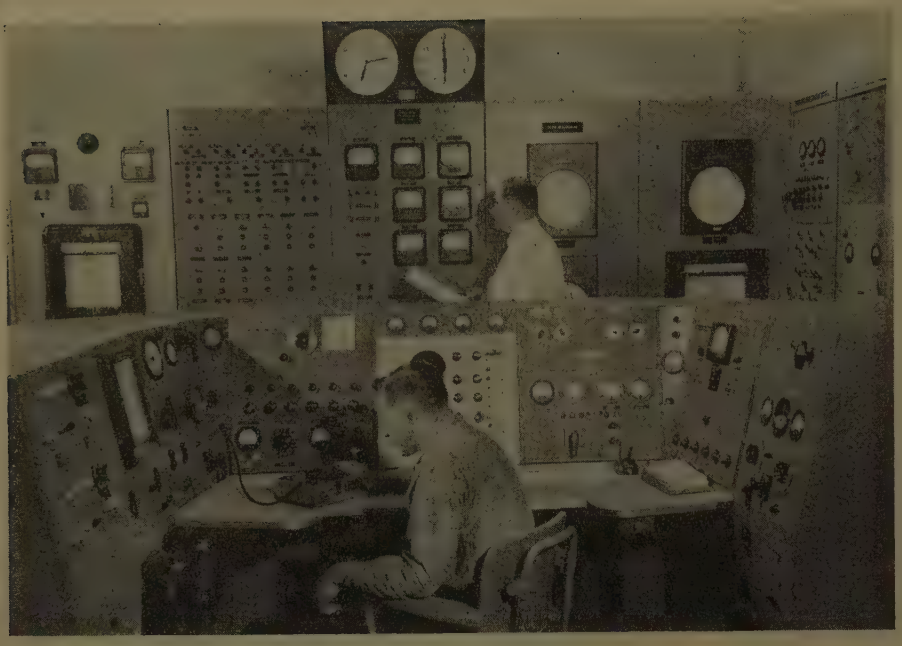


Fig. 5. Cyclotron control room with rectifier control cubicle at upper left

are in the minimum voltage position. As previously discussed, it is desired to energize the magnet at minimum voltage so as to keep eddy currents in the dee structures to a minimum. Magnet current is then adjusted manually to the desired value and then current control is switched to the automatic current regulater. Shutdown procedure is essentially the reverse of startup. Interlocks are provided to prevent an operator from causing faulty operating sequences.

For safety reasons, startup can be accomplished only at the rectifier itself. Operation and shutdown can be done there and remotely from the control room in another building. For remote control a rectifier control board is provided; see Fig. 5. Normal rectifier operation is from the control room.

Two types of automatic shutdown are provided in case of trouble. The first type, the emergency shutdown, drops in the short-circuiting breakers and opens the a-c and d-c breakers. A backfire or any other short-circuit condition will initiate the emergency shutdown.

The second type of automatic shutdown follows the normal manual shutdown sequence. It is initiated by less catastrophic causes than is the emergency shutdown.

Operating Experience

To date no major operating or maintenance problem has arisen. Only three minor troubles have occurred. The first was that the overlap regulator developed a tendency to stick in the physical position corresponding to minimum output voltage. This resulted in a backfire, and the destruction of one contact during shutdown, if the operator failed to notice the sticking. The trouble was cleared up by adding a spring which acts to pull the regulator out of the minimum voltage position.

The second trouble was that the a-c short-circuiting breaker developed a tendency to fall in, thereby causing the rectifier to shut itself down. The cause was simply some dirt between the magnetic latching surfaces. Cleaning the surfaces eliminated the trouble.

The third trouble was a short-circuited insulator in the current recorder-controller which was easily repaired.

Efficiency

To date, all cyclotron operation has been with a 2,000-ampere magnet current at about 260 volts corresponding to about 40% of the rectifier's rated power load. At this load the efficiency is about 94.7%. This has been determined by using an a-c watt-hour meter to determine a-c input, and the d-c recorder and d-c voltmeter for d-c output.

It should be noted that this is an overall rectification efficiency and takes into account the losses in the step-voltage regulator and that most rectifier installations require no step-voltage regulator.

Attendance Necessary

Once turned on, the rectifier requires no attendance, but the oscilloscope and instruments of the rectifier are checked and recorded periodically by the cyclotron operators along with the many other instruments involved in the operation.

Current Regulation

Automatic current control has been more than satisfactory. Specifications called for magnet current to be held automatically to within ± 0.25% of a preset value. At the present 2,000-ampere operating point, the maximum current deviation is ±0.1%, and the average current deviation is ±0.05%.

Procedure for Contact Replacement

To replace a complete set of six contacts and return the rectifier to service takes about 20 minutes and requires no special skill. Timing of the new contacts is checked by running the mechanism in the "test" position and checking a voltmeter for a predetermined reading. Little, if any, timing adjustment is ever required. Timing adjustment is accomplished by rotating an external screw geared to each individual push rod.

With the contacts replaced and the timing checked, the rectifier is ready to deliver load; no "break-in" period is required.

Contact Life

To date contact life has been satisfactory. At the end of 613 hours of operation, the six contacts were removed and inspected for an estimate of contact life. The estimate was 2,000 hours.

Only one backfire has occurred during the cyclotron operation. It was caused by a line voltage disturbance caused by the failure of another feeder in the power company's distribution system. All contacts were damaged and replaced.

Several other contacts have been damaged and replaced due to backfires occurring during the shutdown sequence when operators failed to notice that the overlap regulator was sticking in the minimum voltage position. As previously stated, this trouble has been eliminated.

One contact was replaced after 852 hours of operation when observation of its break via the oscilloscope indicated that it was breaking erratically.

With the present 2,300 operating hours on the rectifier, five of the present contacts have been in use for 1,600 hours and the sixth for 850 hours.

Conclusions

The mechanical rectifier, after 8 months of operation, is proving to be a reliable and efficient power supply for a cyclotron magnet. Automatic current control, to a fine degree, has been obtained. The inherent dangers, due to the high stored energy in the inductance of the magnet, have been surmounted.

Discussion

I. K. Dortort (I-T-E Circuit Breaker Company, Philadelphia, Pa): Mr. Warren has given us a comprehensive and detailed report on the preliminary investigations and selection of a mechanical rectifier for a very special application, the engineering characteristics developed to make the application practical, and the operating experience with this mechanical rectifier supplying power to the magnet coil of a cyclotron having an extremely large stored energy. It is gratifying to have the engineering decisions and calculations of the manfacturer so accurately confirmed.

One of the most difficult and complex problems, and the one involving the greatest risk, involved the starting, stopping, and reversing of current flow from a mechanical rectifier to an extremely inductive load. Specifications restricted the maximum voltage appearing across the cyclotron coil for any reason including inductive voltages, to a value of 500 volts, only 50 volts above rated voltage. A similar restriction was placed on maximum current. These stringent requirements were well justified by the high cost of any repairs to the magnet coil in time and money.

Adding to the difficulty was the desire to shut down, reverse polarity, and start up again in 30 minutes. The natural decay time of the magnet current from 3,000 amperes down to 50 was calculated to be approximately 550 seconds. Considering the fact that below 50 amperes the inductance of the cyclotron coil is above 36 henrys, it was obvious that no switching could be safely attempted at coil currents exceeding a small fraction of an ampere. Obviously, switching arrangements could be devised to insert increasing resistance values as the current decreased, but aside from their complexity such switching arrangements involved a considerable risk of malfunctioning and damage to the cyclotron.

With the scheme finally adopted the estimated time for the decay of current from 3,000 amperes down to 0.6 ampere is 17½ minutes and does not require any switching within the coil circuit. Special selenium rectifier stacks were designed and connected across the coil as a voltage-suppressor valve. The characteristics of the selenium valve were chosen so that the rate of decay of flux in the magnet, interacting with the eddy currents in the conducting parts within the *D*'s, would not produce excessive forces. Aluminum base plates were chosen for low reactance and the physical design balanced for thermal capacity. At 0.6 ampere, it is safe to use the make-before-break reversing switch, and normal startup in the reverse direction can be initiated.

An Analytic Method for Finding the Closed-Loop Frequency Response of Nonlinear Feedback-Control Systems

KATSUHIKO OGATA
ASSOCIATE MEMBER AIEE

THE linear theory in the field of feedback-control systems is highly developed and various techniques are available for analysis and synthesis of linear feedback-control systems. On the other hand, there are many problems to be solved in the field of nonlinear feedback-control systems. Techniques of analyzing and synthesizing nonlinear feedback-control systems are quite limited because of difficulties in handling nonlinear differential equations.

It is well known that frequency-response curves of feedback-control systems with saturation nonlinearities often exhibit the jump-resonance phenomenon. It is also known that responses of nonlinear feedback-control systems sometimes exhibit subharmonic oscillations.

The jump-resonance phenomenon is discussed or explained by authors such as Minorsky,[1] Levinson,[2] Prince, Jr.,[3] and Truxal.[4] Levinson,[2] using an analytical and graphical technique, indicates the multivalued response and the jump-resonance phenomenon of a tachometer-stabilized system. The principle used in his paper is that the effect of saturation is to lower the average-gain value of the path in which saturation takes place. Prince, Jr., using the equivalent-gain method, shows in his paper[3] the jump resonance phenomenon as responses of an acceleration-limited system and an error-limited system.

Subharmonic oscillations as responses of nonlinear vibratory systems are often discussed in the field of nonlinear mechanics. Not much is known, however, about subharmonic oscillations as responses of nonlinear feedback-control systems. When a nonlinear feedback-control system includes saturation, it is seen that at a fixed input frequency and amplitude there may be more than one frequency with which the system output oscillates. For example, when a nonlinear high-gain feedback-control system with saturation is used as a recorder, the output of the recorder may oscillate with a frequency which is $1/n$ (n is an integer greater than one) times the input frequency, if the input of the recorder is periodic with a high frequency. When the frequency of the output oscillation is $1/n$ times that of the input, the output oscillation may be called the $1/n$-order pure subharmonic oscillation. An analysis of pure subharmonic oscillations as responses of feedback-control systems with saturation nonlinearities will be presented by the author in another paper.

The possibility of the occurrence of pure subharmonic oscillations as the response of a nonlinear feedback-control system can be predicted from an accurate closed-loop attenuation-frequency curve for a given input amplitude. For example, when a nonlinear feedback-control system includes saturation, and if the maximum ratio of the system output amplitude to the system input amplitude is large (about 3 decibels or more), there is a possibility that the output of the system may oscillate with a subharmonic of the input frequency. However, if the maximum ratio of the output amplitude to the input amplitude is small (2 decibels or less), the possibility of the occurrence of pure subharmonic oscillations is very small or nonexistent.

The output waveforms of nonlinear elements are often extremely nonsinusoidal. Because of the difficulty in handling such nonsinusoidal waveforms, the well-known describing-function method[5,6] which is essentially a graphical method takes only the fundamental-harmonic component into consideration

Paper 57-634, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24–28, 1957. Manuscript submitted October 23, 1956; made available for printing April 16, 1957.

KATSUHIKO OGATA is with the University of Minnesota, Minneapolis, Minn.

The material in this paper is based on a part of a dissertation submitted by the author in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Engineering Science at the University of California. The author wishes to acknowledge with sincere gratitude many discussions offered during the course of the research by Prof. A. M. Hopkin who acted as chairman of the Dissertation Committee.

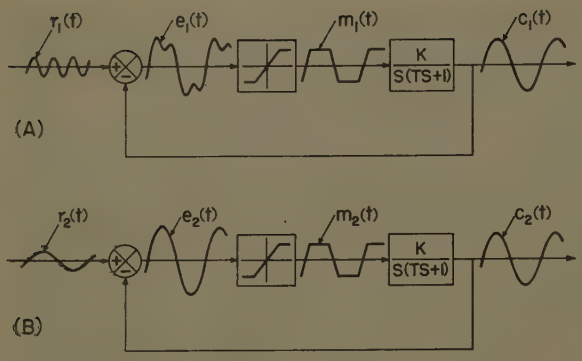


Fig. 1. Block diagrams of second-order feedback-control system with saturation nonlinearity showing

A—Pure subharmonic oscillation

B—Fundamental oscillation

for analyzing and synthesizing nonlinear feedback-control systems. To the author's knowledge, no analytic methods for finding closed-loop frequency-response characteristics of nonlinear feedback-control systems and/or jump-resonance frequencies are found in past publications.

The main purpose of this paper is to present an analytic method for finding the accurate closed-loop frequency-response characteristics of nonlinear feedback-control systems. By means of the periodic-response function, defined herein as the Laplace transform of the steady-state waveform of the output of a nonlinear element when the input to that element is periodic, this analytic method takes all or most of the harmonic components of the output of the nonlinear element into consideration for finding the closed-loop frequency-response characteristics. Accurate closed-loop frequency-response results obtained by this method may be used for predicting the possibility of the occurrence of pure subharmonic oscillations. In this paper methods are also presented for finding the jump-resonance frequencies in terms of the system constants and the input amplitude.

The emphasis of this paper is placed on showing the general approach of the method. Therefore, nonlinear second-order feedback-control systems are used as examples. However, the method presented is applicable to nonlinear higher order feedback-control systems. An application of this method for finding the closed-loop frequency-response characteristics of a nonlinear higher order feedback-control system has been shown previously.[7]

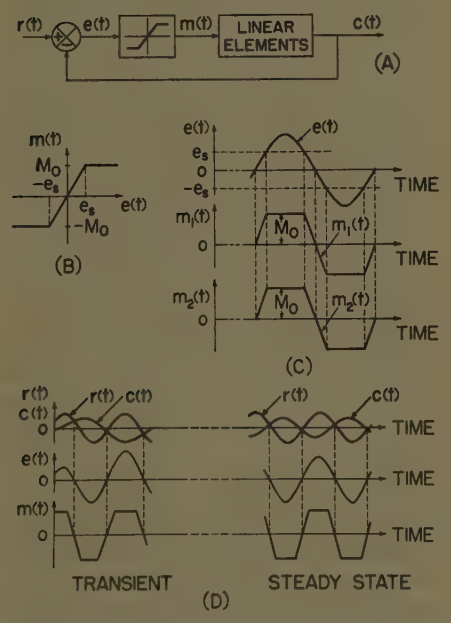


Fig. 2. Block diagram, nonlinearity, waveforms c(t), e(t), and m(t) of nonlinear feedback-control system

Differences Between the Method Presented and the Describing-Function Method

Some differences between the method presented in this paper and the describing-function method are listed in the following.

1. The author's method takes into account the effects of all or most of the higher harmonics of the nonsinusoidal output of the nonlinear element for finding the system output, while the describing-function method takes into consideration only the fundamental harmonic of the nonsinusoidal output of the nonlinear element.

2. The author's method is essentially analytical, while the describing-function method is essentially graphical.

3. The author's method does not require the assumption that the input to the nonlinear element is sinusoidal. On the contrary, it is the fundamental assumption of the describing-function method that the input to the nonlinear element is sinusoidal; therefore, the method presented in this paper is valid for finding closed-loop periodic-response characteristics of nonlinear feedback-control systems for periodic inputs such as sinusoidal inputs, square-wave inputs, triangular inputs, etc., while the describing-function method is valid only for sinusoidal inputs.

4. The author's method is valid for finding the pure subharmonic output of the system. Fig. 1(A) shows a pure subharmonic oscillation of the output of a feedback-control system with a saturation nonlinearity. Fig. 1(B) shows a fundamental oscillation of the output of the same system. The steady-state waveform of $e_1(t)$ is quite different from a sinusoidal wave, therefore, the describing-function method is not valid when the system output exhibits the pure subharmonic oscillation as shown in Fig. 1(A). However, the method presented in this paper is valid, because the waveform of $m_1(t)$ is used for finding $c_1(t)$ analytically.

Analytic Method for Finding Closed-Loop Frequency Response of Nonlinear Feedback-Control Systems

Periodic-Response Function

Fig. 2(A) shows the block diagram of a feedback-control system with a saturation nonlinearity. The input-output characteristic of the nonlinear element is shown in Fig. 2(B). The exact output waveform of the nonlinear element may be shown as the curve $m_1(t)$ in Fig. 2(C). The waveform $m_2(t)$ is a straight-line approximation of $m_1(t)$. The error signal $e(t)$ and the waveform $m_2(t)$ are also shown in Fig. 2(C). The approximate waveform $m_2(t)$ may be the same as the exact waveform $m_1(t)$, when the latter has only straight-line segments, e.g., the output waveform of an on-off nonlinearity. After a sinusoidal input is applied to the system and steady state has been reached, the output of the nonlinear element will be periodic. Let the output waveform of the nonlinear element be $m(t)$. Fig. 2(D) shows the transient and steady-state waveforms $e(t)$, $m(t)$, and $c(t)$ together with the input $r(t)$.

By defining $M(s)$ which is the Laplace transform of $m(t)$, the steady-state waveform of the output of the nonlinear element when the input to that element is periodic, as the "periodic-response function," it may be written as follows,[8] if the period of the waveform $m(t)$ is $2a$.

$$M(s)=\frac{\int_0^{2a}\epsilon^{-st}m(t)dt}{1-\epsilon^{-2as}} \qquad (1)$$

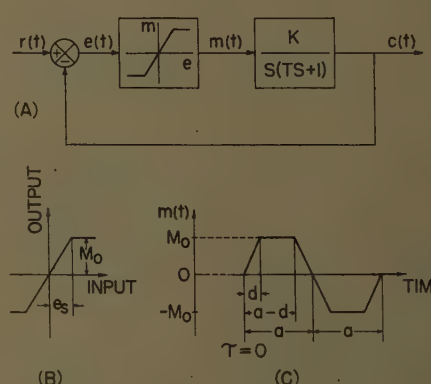


Fig. 3. A—Block diagram of second-order feedback-control system with saturation nonlinearity. B—Input-output characteristic of nonlinearity. C—Waveform m(t)

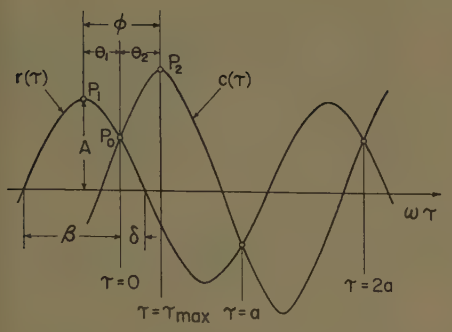


Fig. 4. Definition of phase angle ϕ

Example: Periodic-Response Function of Saturation Nonlinearity. Fig. 3(A) shows the block diagram of a second-order feedback-control system with a saturation nonlinearity. The input-output characteristic of the nonlinearity is shown in Fig. 3(B). The steady-state waveform of the signal $m(t)$ is shown in Fig. 3(C). When the degree of saturation is relatively high, the waveform of the output of the saturation nonlinearity is trapezoidal. The periodic-response function of the saturation nonlinearity may be obtained as a function of M_o, a, and d, if M_o and d are defined as follows: M_o is the maximum output of the nonlinearity, and d is the time interval when the output changes from zero to the point where saturation starts as shown in Fig. 3(C). In general, d is a function of a, the half-period of the output, and E, the amplitude of the error signal. Using equation 1, the periodic-response function of the saturation nonlinearity, the Laplace transform of the waveform $m(t)$ shown in Fig. 3(C) may be written as follows

$$M(s)=\frac{M_o}{ds^2}\times\left[1-(\epsilon^{-ds}+\epsilon^{ds-as})\sum_{n=0}^{\infty}(-1)^n\epsilon^{-ans}\right] \tag{2}$$

Method of Approach

The analytic method presented in this paper uses the following procedure:

1. Find the exact or the approximate waveform $m(t)$, the steady-state output of a nonlinear element.
2. Find $V(s)$, the Laplace transform of the system output velocity $v(t)$, in terms of the periodic-response function $M(s)$, the Laplace transform of $m(t)$, and the system transfer functions.
3. Find the inverse Laplace transform of $V(s)$ and obtain the output velocity as a function of the output frequency.
4. Integrate the output velocity with respect to time and find the system output $c(t)$ as a function of the output frequency.

Obtain the relation between the output amplitude and the output frequency; also obtain the relation between the output phase lag and the output frequency.

Closed-Loop Attenuation-Frequency Characteristics

When the system input $r(t)$ is sinusoidal i.e., $r(t)=A\sin(\omega t+\alpha)$, the attenuation-frequency characteristics for the second-order feedback-control system with a saturation nonlinearity shown in Fig. 3(A) may be found by obtaining the equation expressing the ratio of the output amplitude C_p to the input amplitude A, as a function of the output frequency. In Fig. 3(A), $c(t)$ and $m(t)$ are related by the following equation

$$(Tp^2+p)c(t)=Km(t) \tag{3}$$

where $p=d/dt$. Taking the time origin $t=0$ so that $r(0)=c(0)$, $e(0<t<a)>0$, and using equations 2 and 3, $V(s)$, the Laplace transform of $v(t)$ where $v(t)=pc(t)$, may be written as follows

$$V(s)=\frac{KM_o}{d}\left(\frac{1}{s^2}-\frac{T}{s}+\frac{T}{s+\frac{1}{T}}\right)\times\left[1-(\epsilon^{-ds}+\epsilon^{ds-as})\sum_{n=0}^{\infty}(-1)^n\epsilon^{-ans}\right]+\frac{v(0)}{s+\frac{1}{T}} \tag{4}$$

where $v(0)$ is the initial velocity. Equation 4 shows the Laplace transform of the velocity $v(t)$ of the system output when the input to the linear element, whose transfer function is $K/s(Ts+1)$, has a trapezoidal waveform as shown in Fig. 3(C).

It should be noted here that in the case of the nonlinear second-order feedback-control system, the classical method to solve equation 3 in time domain may be simpler. The purpose of using the Laplace-transform approach shown here

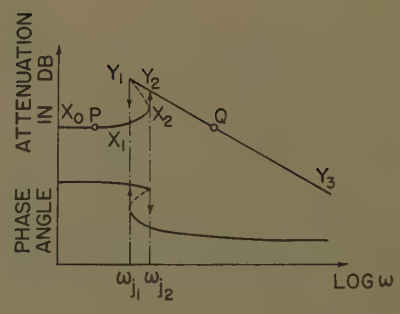


Fig. 5. Attenuation-frequency and phase-frequency curves showing jump resonance

and in the appendix is to describe the procedure equally applicable to nonlinear higher order feedback-control systems. When the linear elements of the nonlinear feedback-control system have double or triple poles, it will be difficult to find periodic solutions by working in the time domain.

Taking the inverse Laplace transform of equation 4 for the time interval $\lambda a\leqslant t\leqslant(\lambda+1)\,a$, where a is the half-period of the output and λ is a sufficiently large odd number, letting $t-\lambda a=\tau$, and integrating $v(\tau)$ with respect to τ, the system output $c(\tau)$ may be obtained as a function of a and τ. If the velocity of the system output becomes zero at some instant during the time interval d and $a-d$, shown in Fig. 3(C), the maximum of $c(\tau)$ for a half-cycle $(0\leqslant\tau\leqslant a)$ occurs during the time interval $d\leqslant\tau\leqslant a-d$. Let τ_{max} be the time when the output position during a half-cycle $(0\leqslant\tau\leqslant a)$ becomes maximum. Integrating $v(d\leqslant\tau\leqslant a-d)$ with respect to τ and substituting $\tau=\tau_{max}$, the output amplitude C_p may be obtained as follows:

$$C_p=KM_oT\left(\ln\cosh\frac{a}{2T}-\ln\sinh\frac{d}{T}+\ln\frac{d}{T}\right) \tag{5}$$

Equation 5 is the analytic solution for finding the system output amplitude C_p of the system shown in Fig. 3(A) as a function of a and d. In equation 5 d is a

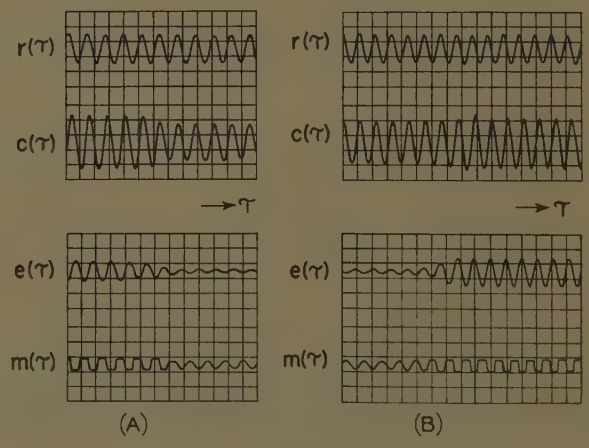


Fig. 6. Curves $r(\tau)$, $c(\tau)$, $e(\tau)$, and $m(\tau)$

A—Showing lower jump resonance

B—Showing upper jump resonance

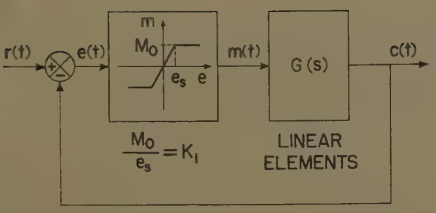


Fig. 7. Block diagram of nonlinear feedback-control system with saturation

function of a and E, the amplitude of the error signal, which is a function of the amplitude and frequency of the input. Substituting $a=\pi/\omega$ in equation 5 and dividing both sides of the equation by A, the following equation may be obtained:

$$\frac{C_p}{A}=\frac{KM_oT}{A}\times\left(\ln\cosh\frac{\pi}{2T\omega}-\ln\sinh\frac{d}{T}+\ln\frac{d}{T}\right) \quad (6)$$

The attenuation-frequency curve of the system shown in Fig. 3(A) under nonlinear operation may be obtained from equation 6.

Closed-Loop Phase-Frequency Characteristics

Since the waveform of the system output is not sinusoidal for nonlinear operation, the phase difference between the input and the output is defined in this paper as the phase difference between the successive maxima of the system input and the system output.

The phase lag ϕ is shown in Fig. 4. In Fig. 4 the input is sinusoidal, i.e., $r(\tau)=A\sin(\omega\tau+\beta)$, and β is the initial phase angle of the input with respect to τ. As may be seen from Fig. 4, the phase lag ϕ is the sum of θ_1 and θ_2, where θ_1 is the phase difference between P_0, the point which corresponds to $\tau=0$, and P_1, the point which corresponds to the maximum of the system input $r(\tau)$; and θ_2 is the phase difference between P_0 and P_2, the point which corresponds to the maximum of the system output $c(\tau)$ for a particular output frequency. Since $c(0)$, the system output magnitude at $\tau=0$, equals $A\sin\beta$; θ_1 is the difference between β and 90 degrees; and θ_2 equals $\omega\tau_{max}$, the phase lag ϕ of the output of the system shown in Fig. 3(A) may be written as a function of ω, the system output frequency, and d as follows:

$$\phi=180^\circ\left[1-\frac{T\omega}{\pi}\left(\ln\cosh\frac{\pi}{2T\omega}+\ln\frac{d}{T}-\ln\sinh\frac{d}{T}\right)\right]-\delta \quad (7)$$

where

$$\sin\delta=\frac{KM_oT}{Ad}\times\left[-T+T\left(\cosh\frac{d}{T}-\sinh\frac{d}{T}\tanh\frac{\pi}{2T\omega}\right)+\frac{\left(\frac{\pi}{\omega}-d\right)d}{2T}\right],\ 0^\circ\leqslant\delta\leqslant90^\circ \quad (8)$$

and the phase angle ϕ is obtained in degrees. Angle δ is indicated in Fig. 4.

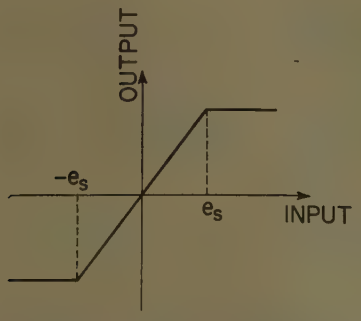


Fig. 8. Input-output characteristic of a nonlinearity showing the abrupt changes at $e=e_s$ and $e=-e_s$

Jump-Resonance Frequencies

The response of a nonlinear feedback-control system with saturation is multivalued when the system input lies within a certain range of amplitude and frequency. As shown in Fig. 5, if a sinusoidal input of a low frequency is fed to a nonlinear feedback-control system with saturation and this frequency is gradually increased, the point P on the attenuation-frequency curve representing the output amplitude of the system follows the curve $X_0X_1X_2$. When the input frequency equals ω_{j2}, the amplitude and the phase of the output change abruptly as shown in Fig. 5. With a further increase in the input frequency, the point Q on the attenuation-frequency curve follows the curve Y_2Y_3. If the procedure is reversed and the input frequency is decreased gradually from a higher value, the point Q follows the curve $Y_3Y_2Y_1$. When the input frequency is decreased further and equals ω_{j1}, the abrupt changes in the output amplitude and phase occur. Calling the frequencies corresponding to ω_{j1} and ω_{j2} the lower jump-resonance frequency and the upper jump-resonance frequency respectively, they may be found as shown in the following.

Lower Jump-Resonance Frequency ω_{j1}

When a feedback-control system has a saturation nonlinearity, at a fixed input frequency and amplitude, there may be more than one $m(t)$ signal which satisfies the loop. In Fig. 5, when the frequency of the input lies between ω_{j1} and ω_{j2}, there can be two stable waveforms of $m(t)$ signal. These are a trapezoidal waveform corresponding to a nonlinear operation of the system and a sinusoidal waveform corresponding to a linear operation of the system. In Fig. 5 curve $X_0X_1X_2$ corresponds to a linear operation and curve $Y_1Y_2Y_3$ corresponds to a nonlinear operation of the system. When a system has a saturation nonlinearity, the magnitude of the system output is always limited. The lower jump-resonance frequency ω_{j1} corresponds to the largest output magnitude and is the lowest input frequency for which the system may be operated nonlinearly.

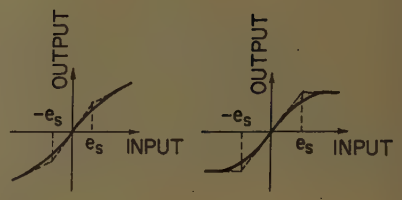


Fig. 9. Approximate values of e_s by bilinear approximation

At the intersection of curves $r(\tau)$ and $c(\tau)$ at $\tau=0$ in Fig. 4,

$$c(0)=A\sin\beta \quad (9)$$

Analytically, the output $c(0)$ of the system shown in the Fig. 3 may be written as a function of a, the half-period of the output, and d as follows:

$$c(0)=\frac{KM_oT}{d}\times\left[-T+T\left(\cosh\frac{d}{T}-\sinh\frac{d}{T}\times\tanh\frac{a}{2T}\right)\right]+\frac{KM_o}{2}(a-d) \quad (10)$$

The detail of the derivation of equation 10

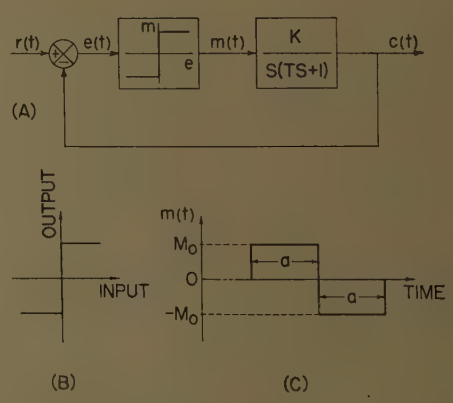


Fig. 10. A—Block diagram of second-order feedback-control system with on-off nonlinearity. B—Input-output characteristic of nonlinearity. C—Waveform m(t)

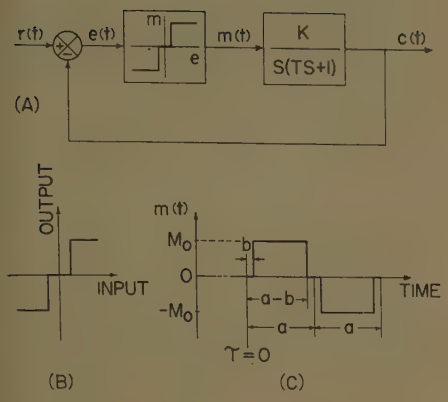


Fig. 11. A—Block diagram of second-order feedback-control system with on-off and dead-zone nonlinearity. B—Input-output characteristic of nonlinearity. C—Waveform m(t)

is shown in the appendix. For the system shown in Fig. 3, it can be operated nonlinearly for the region 90 degrees $\leqslant\beta\leqslant 180$ degrees. Examination of equation 10 indicates that $c(0)$ becomes large as the value of a, the half-period of the output, increases, and vice versa. For the region $\beta<90$ degrees, the nonlinear operation is unstable. When β becomes slightly less than 90 degrees, the value of $c(0)$ becomes smaller than that of $c(0)$ corresponding to $\beta=90$ degrees, with the result that the amplitude and phase lag of the output become smaller than those of the output corresponding to $\beta=90$ degrees. The decrease in the amplitude and phase lag of the output makes the amplitude of the error signal smaller than that of the error signal corresponding to $\beta=90$ degrees. This causes the resulting amplitude of the output to be smaller than the preceding one. The phase lag of the output also becomes smaller than the preceding value. This process continues until the steady state is reached. For the system shown in Fig. 3, the steady state corresponds to the state where the system output oscillates in a linear fashion. The process may be seen from curve $c(\tau)$, the system output; curve $e(\tau)$, the error signal; and the curve $m(\tau)$, the output of the nonlinear element, as shown in Fig. 6(A). It is clear from equation 5 that the amplitude of the system output under nonlinear operation becomes large as the value of a, the half-period of the output, increases. Therefore, C_p becomes maximum when $c(0)$ becomes maximum. From equation 9, $c(0)$ becomes maximum when β equals 90 degrees. Let the value of a, the half-period of the output, which satisfies the equation

$$c(0)=A \tag{11}$$

be a_1. Then the lower jump-resonance frequency ω_{j1} may be obtained as the following equation

$$\omega_{j1}=\frac{\pi}{a_1}. \tag{12}$$

Upper Jump-Resonance Frequency ω_{j2}

When the frequency of the input to the system with a saturation nonlinearity shown in Fig. 7 is low, the amplitude of the error signal is small, with the result that the system is operated linearly. When the frequency of the system input is increased from a low value, at a certain input frequency the amplitude of the error signal equals e_s, where e_s is the value of the input to the nonlinear element at which the input-output relation of the nonlinearity changes abruptly as shown in Fig. 8. When the amplitude of the error signal exceeds e_s, the magnitude of $m(\tau)$ (the output of the nonlinear element) does not increase as much as it should if the system were linear; the result is that the amplitude of the error signal for the following cycle becomes larger than before as does the phase lag of $c(\tau)$, the output of the system, for the following cycle, until the equilibrium state is reached. This equilibrium state corresponds to the nonlinear operation of the system and the waveform $m(\tau)$ is trapezoidal. The upper jump resonance may be seen from curve $c(\tau)$ shown in Fig. 6(B). Curves $e(\tau)$ and $m(\tau)$ at the upper jump resonance are also shown in Fig. 6(B). As may be seen from the curves in Fig. 6(B), the upper jump resonance takes place when the amplitude of the error signal equals e_s.

When there is no abrupt change in the input-output characteristic of the nonlinearity, the approximate value of e_s may be found by the bilinear approximation as shown in Fig. 9. In the cases of the nonlinearities shown in Fig. 9, the changes in the shapes of curves $e(\tau)$ and $m(\tau)$ take place gradually when the amplitude of the error signal $e(\tau)$ approaches e_s from a smaller value. When the amplitude of the error signal $e(\tau)$ approximately equals e_s, the upper jump resonance takes place.

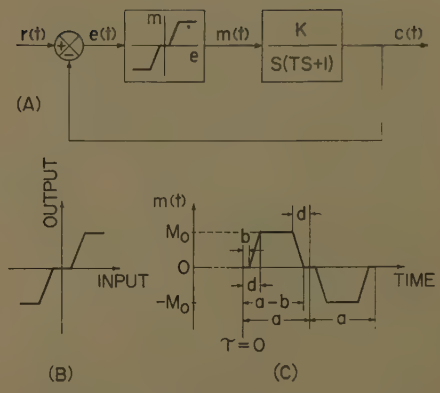


Fig. 12. A—Block diagram of second-order feedback-control system with saturation and dead-zone nonlinearity. B—Input-output characteristic of nonlinearity. C—Waveform m(t)

Since the limit of the linear operation of the system corresponds to the frequency at which the magnitude of the error signal equals e_s, the upper jump-resonance frequency ω_{j2} of the output of the system shown in Fig. 7 may be found to be frequency ω satisfying the following equation

$$\left|\frac{A}{1+K_1G(j\omega)}\right|=e_s \tag{13}$$

In equation 13, K_1 is the linear gain of the saturation nonlinearity, G is the transfer function of the linear elements, and A is the amplitude of the input sinusoid.

Analytic Frequency-Response Solutions of Nonlinear Second-Order Feedback-Control Systems

Second-Order Feedback-Control System with Saturation Nonlinearity

The analytic frequency-response solution of the system shown in Fig. 3(A) under nonlinear operation may be obtained from equations 6 through 8, if the maximum of the system output for a half-cycle ($0\leqslant\tau\leqslant a$) at steady state occurs during the time interval $d\leqslant\tau\leqslant a-d$, shown in Fig. 3(C). Since the system may be operated linearly if the input frequency is

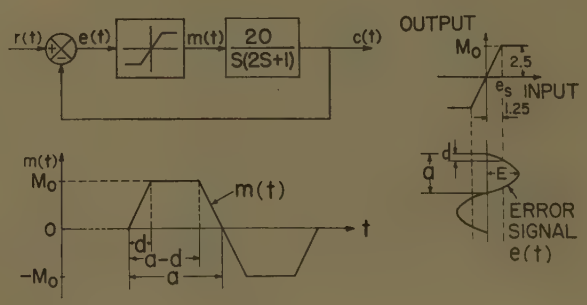


Fig. 13. Block diagram, nonlinearity, curves e(t) and m(t) of second-order feedback-control system with saturation nonlinearity

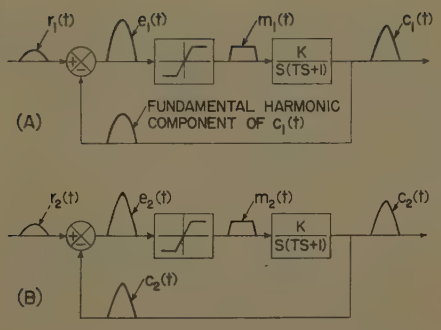


Fig. 14. Waveforms of various signals of nonlinear feedback-control system with saturation

A—When feedback signal is assumed to be fundamental-harmonic component of output
B—When feedback signal is exact output

not higher than the upper jump-resonance frequency ω_{j2}, the linear theory is used to find the attenuation-frequency and phase-frequency characteristics for the system under linear operation. The lower jump-resonance frequency ω_{j1} may be found by substituting equation 10 and the relation $a_1=\pi/\omega_{j1}$ into equation 11 in the following manner

$$\frac{T}{d}\left(\cosh\frac{d}{T}-\sinh\frac{d}{T}\times\tanh\frac{\pi}{2T\omega_{j1}}-1\right)+\frac{\pi}{2T\omega_{j1}}-\frac{d}{2T}=\frac{A}{KM_oT} \quad (14)$$

where ω_{j1} is obtained in rad/sec (radians per second).

The upper jump-resonance frequency ω_{j2} may be found from equation 13 as follows:

$$\left|\frac{Aj\omega_{j2}(Tj\omega_{j2}+1)}{j\omega_{j2}(Tj\omega_{j2}+1)+M_oK/e_s}\right|=e_s \quad (15)$$

When d is small compared with a, the half-period of the output, the approximation by setting $d=0$ may be suggested in order to simplify the numerical calculation unless a very high accuracy is needed. When d is set to be zero, equations 16 through 19 should be used instead of equations 6 through 8 and 14.

Second-Order Feedback-Control System with On-Off Nonlinearity

Fig. 10(A) shows the block diagram of a second-order feedback-control system with an on-off nonlinearity. Fig. 10(B) shows the input-output characteristic of the nonlinearity. The waveform of $m(t)$ is shown in Fig. 10(C). The closed-loop frequency-response curves corresponding to the inputs having frequencies which are equal to or higher than the lower jump-resonance frequency for a particular input amplitude may be obtained from the three equations which follow

$$\frac{C_p}{A}=\frac{KM_oT}{A}\ln\cosh\frac{\pi}{2T\omega} \quad (16)$$

$$\phi=180°\left(1-\frac{T\omega}{\pi}\ln\cosh\frac{\pi}{2T\omega}\right)-\sin^{-1}\frac{KM_oT}{A}\left(\frac{\pi}{2T\omega}-\tanh\frac{\pi}{2T\omega}\right) \quad (17)$$

where

$$0°\leqslant\sin^{-1}\left[\frac{KM_oT}{A}\left(\frac{\pi}{2T\omega}-\tanh\frac{\pi}{2T\omega}\right)\right]\leqslant 90° \quad (18)$$

and the phase lag ϕ is obtained in degrees. The lower jump-resonance frequency ω_{j1} may be obtained from the following equation

$$\frac{\pi}{2T\omega_{j1}}-\tanh\frac{\pi}{2T\omega_{j1}}=\frac{A}{KM_oT} \quad (19)$$

where ω_{j1} is obtained in rad/sec.

Second-Order Feedback-Control System with On-Off and Dead-Zone Nonlinearity

The block diagram of a second-order feedback-control system with an on-off and dead-zone nonlinearity is shown in Fig. 11(A). The input-output characteristic and the output waveform $m(t)$ of the nonlinearity are shown in Figs. 11(B) and (C) respectively. When the maximum of the system output during the half-cycle ($0\leqslant\tau\leqslant a$) at the steady state occurs during the time interval $b\leqslant\tau\leqslant a-b$, where b is the time corresponding to the dead-zone amplitude as shown in Fig. 11(C), the closed-loop frequency-response curves corresponding to the inputs having frequencies which are equal to or higher than the lower jump-resonance frequency for a particular input amplitude may be obtained from the following equations

$$\frac{C_p}{A}=\frac{KM_oT}{A}\left(\ln\cosh\frac{\pi}{2T\omega}-\ln\cosh\frac{b}{T}\right) \quad (20)$$

$$\phi=180°\left[1-\frac{T}{a}\left(\ln\cosh\frac{\pi}{2T\omega}-\ln\cosh\frac{b}{T}\right)\right]-\sin^{-1}\left[\frac{KM_oT}{A}\left(\frac{\pi}{2T\omega}-\frac{b}{T}-\cosh\frac{b}{T}\times\tanh\frac{\pi}{2T\omega}+\sinh\frac{b}{T}\right)\right] \quad (21)$$

where

$$0°\leqslant\sin^{-1}\left[\frac{KM_oT}{A}\left(\frac{\pi}{2T\omega}-\frac{b}{T}-\cosh\frac{b}{T}\times\tanh\frac{\pi}{2T\omega}+\sinh\frac{b}{T}\right)\right]\leqslant 90° \quad (22)$$

and the phase lag ϕ is obtained in degrees. The lower jump-resonance frequency ω_{j1} may be obtained from the following:

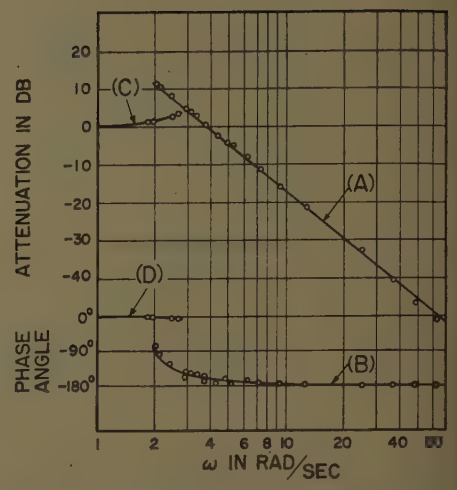


Fig. 15. Attenuation-frequency and phase-frequency characteristics of system shown in Fig. 13

$$\frac{\pi}{2T\omega_{j1}}-\cosh\frac{b}{T}\times\tanh\frac{\pi}{2T\omega_{j1}}-\frac{b}{T}+\sinh\frac{b}{T}=\frac{A}{KM_oT} \quad (23)$$

Second-Order Feedback-Control System with Saturation and Dead-Zone Nonlinearity

The block diagram of a second-order feedback-control system with a saturation and dead-zone nonlinearity is shown in Fig. 12(A). Figs. 12(B) and (C) show the input-output characteristic and the output waveform of the nonlinearity. When the maximum of the system output during the half-cycle ($0\leqslant\tau\leqslant a$) at the steady state occurs during the time interval $d\leqslant\tau\leqslant a-d$, shown in Fig. 12(C), the closed-loop frequency-response curves corresponding to the inputs having frequencies which are equal to or higher than the lower jump-resonance frequency for a particular input amplitude may be obtained from the following equations

$$\frac{C_p}{A}=\frac{KM_oT}{A}\left[\ln\frac{d-b}{T}+\ln\cosh\frac{\pi}{2T\omega}-\ln\left(\sinh\frac{d}{T}-\sinh\frac{b}{T}\right)\right] \quad (24)$$

$$\phi=180°\left\{1-\frac{T\omega}{\pi}\left[\ln\frac{d-b}{T}+\ln\cosh\frac{\pi}{2T\omega}-\ln\left(\sinh\frac{d}{T}-\sinh\frac{b}{T}\right)\right]\right\}-\delta \quad (25)$$

where

$$\delta=\sin^{-1}\left(\frac{KM_oT}{A}\left\{\frac{\pi}{2T\omega}-\frac{d+b}{2T}-\frac{T}{d-b}\times\left[\cosh\frac{b}{T}-\cosh\frac{d}{T}+\tanh\frac{\pi}{2T\omega}\times\left(\sinh\frac{d}{T}-\sinh\frac{b}{T}\right)\right]\right\}\right),\ 0°\leqslant\delta\leqslant 90° \quad (26)$$

and the phase lag ϕ is obtained in degrees.

The lower jump-resonance frequency ω_{j1} may be obtained from the following:

$$\frac{\pi}{2T\omega_{j1}}-\frac{T}{d-b}\tanh\frac{\pi}{2T\omega_{j1}}\times\left(\sinh\frac{d}{T}-\sinh\frac{b}{T}\right)-\frac{d+b}{2T}-\frac{T}{d-b}\left(\cosh\frac{b}{T}-\cosh\frac{d}{T}\right)=\frac{A}{KM_oT} \quad (27)$$

where ω_{j1} is obtained in rad/sec.

Example 1. Frequency-Response Solutions of Feedback-Control System Shown in Fig. 13.

Approximate Solution. Assume that d is equal to zero. In this example $K=20$, $T=2$ sec, $M_o=2.5$, and $e_s=1.25$. The system input $r(t)$ is assumed to be a sinusoid $A\sin\omega t$, where $A=2$. Since the equations giving the analytic frequency-response solutions should be used for the inputs having frequencies equal to or higher than the lower jump-resonance frequency corresponding to the particular input amplitude, the lower jump-resonance frequency is found first. By substituting numerical values of A, K, T, and M_o into equation 19, the lower jump-resonance frequency ω_{j1} may be found from the following equation

$$\frac{0.785}{\omega_{j1}}-\tanh\frac{0.785}{\omega_{j1}}=0.02 \quad (28)$$

By using a mathematical table, it may be found that $\tanh\chi$ is equal to 0.37995 when χ is equal to 0.40. Therefore, when

$$\frac{0.785}{\omega_{j1}}=0.40 \quad (29)$$

equation 28 is almost satisfied. From equation 29

$$\omega_{j1}=1.96 \text{ rad/sec} \quad (30)$$

When the input frequency is equal to or higher than 1.96 rad/sec, the attenuation-frequency and phase-frequency characteristics may be obtained from equations 16 through 18. The upper jump-resonance frequency may be found by substituting the numerical values of A, K, T, M_o, and e_s into equation 15. The upper jump-resonance frequency ω_{j2} thus found is 2.77 rad/sec. When the input frequency is equal to or less than 2.77 rad/sec, the attenuation-frequency and phase-frequency characteristics may be obtained from the linear theory.

Accurate Solution. Equations 6 through 8 should be used for finding accurate attenuation-frequency characteristics and phase-frequency characteristics. It should be noted here that when the analytic method presented in this paper is used for finding frequency-response characteristics, the feedback signal can be either sinusoidal or nonsinusoidal. Fig. 14(A) shows the waveforms of the signals $r_1(t)$, $e_1(t)$, $m_1(t)$, and $c_1(t)$ when the feedback signal is the fundamental-harmonic component of $c_1(t)$. Fig. 14(B) shows the waveforms of the signals $r_2(t)$, $e_2(t)$, $m_2(t)$, and $c_2(t)$ when the exact output $c_2(t)$ is fed back. The waveforms of $e_1(t)$ and $e_2(t)$ are different from each other, but the waveforms $m_1(t)$ and $m_2(t)$ are almost the same. Therefore, to simplify the numerical calculation, the feedback signal may be assumed to be sinusoidal. The amplitude of the feedback sinusoid may be assumed to be either the amplitude C_p of the system output or the amplitude of the fundamental-harmonic component of the system output.

Since equations 6 through 8 are functions of ω and d which is a function of a and E, the amplitude of the error signal; the values of E for various values of a or ω must be found first. Assume that the feedback signal is a sinusoid with an amplitude of C_p. Then the following equation may be obtained from Fig. 13

$$d=\frac{a}{\pi}\sin^{-1}\left(\frac{e_s}{E}\right) \quad (31)$$

where $a/\pi=1/\omega$

By simple analysis the following equation may also be obtained

$$E=\sqrt{A^2-2AC_p\cos\phi+C_p^2} \quad (32)$$

where ϕ is the output phase lag. In equation 32, C_p and ϕ may be obtained from equations 16 through 18. From equations 31 and 32, the values of d for various values of a or ω may be found. Then the accurate attenuation-frequency and phase-frequency characteristics for the system under nonlinear operation may be found from equations 6 through 8 by using the values of d thus obtained. The accurate lower jump-resonance frequency ω_{j1} may be found from equation 14. The value of ω_{j1} obtained from equation 14 is 1.99 rad/sec. The attenuation-frequency and phase-frequency curves obtained by use of equations 6 through 8 are shown in Fig. 15 as curves (A) and (B) respectively. The attenuation-frequency and phase-frequency curves corresponding to the linear operation of the system are shown in Fig. 15 as curves (C) and (D) respectively. The results obtained by the analytic method were compared with those obtained by an electronic-analog-computer study. The values of the lower jump-resonance frequency and the upper jump-resonance frequency obtained by the analytic method are 1.99 rad/sec and 2.77 rad/sec respectively, and those obtained by experiments are 2.01 rad/sec and 2.75 rad/sec respectively.

The experimental results obtained by use of an electronic-analog-computer are plotted in Fig. 15 as circles. Fig. 15 shows that the theoretical closed-loop attenuation-frequency and phase-frequency curves agree with the experimental results satisfactorily.

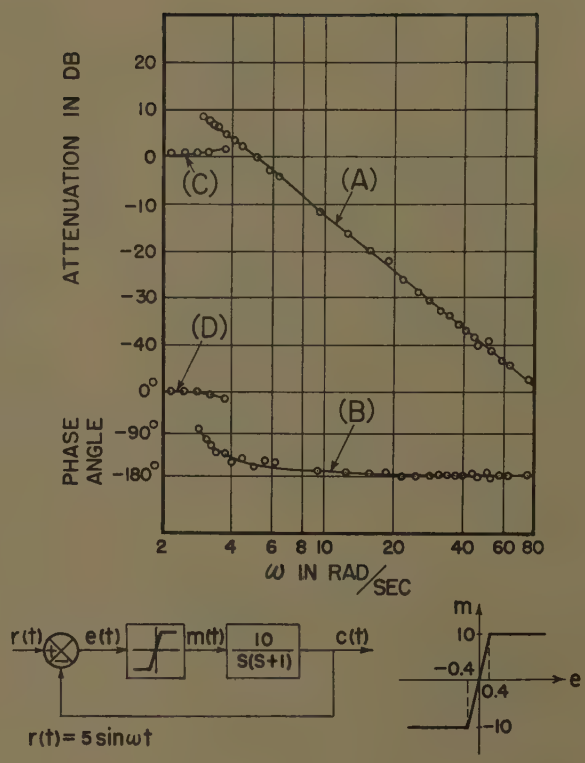


Fig. 16. Attenuation-frequency and phase-frequency characteristics of second-order feedback-control system with saturation nonlinearity

Example 2. Frequency-Response Solution of Feedback-Control System Shown in Fig. 16.

The numerical values for the system constants are as follows: $K=10$, $M_o=10$, $T=1$ sec, and $e_s=0.4$. The amplitude A of the input sinusoid is five.

The attenuation-frequency and phase-frequency curves of the system under nonlinear operation are shown in Fig. 16 as curves (A) and (B) respectively. Curves (C) and (D) in Fig. 16 correspond to the linear operation of the system. The experimental values obtained by means of an electronic analog computer are plotted as circles in Fig. 16.

Conclusions

1. An analytic method for finding the accurate closed-loop attenuation-frequency and phase-frequency characteristics for the nonlinear control systems by means of the periodic-response function is presented. The method takes into account the effects of all or most of the higher harmonic components of the nonsinusoidal output of a nonlinear element for finding the system output.

2. The analytic closed-loop frequency-response solutions for fundamental oscillations are shown for the following systems:
   a. The second-order feedback-control system with the saturation nonlinearity shown in Fig. 3.
   b. The second-order feedback-control system with the on-off nonlinearity shown in Fig. 10.
   c. The second-order feedback-control system with the on-off and dead-zone nonlinearity shown in Fig. 11.
   d. The second-order feedback-control system with the saturation and dead-zone nonlinearity shown in Fig. 12.

The closed-loop attenuation-frequency and phase-frequency characteristics for the foregoing systems for fundamental oscillations may be obtained by merely substituting system constants and input amplitudes into corresponding equations shown in the paper. Since the equations developed in this paper giving amplitudes and phase lags of the system output as functions of the output frequency contain the system constants as parameters, the contribution of these constants to the attenuation-frequency and phase-frequency characteristics is evident. Therefore, these equations may be used for synthesizing corresponding nonlinear feedback-control systems.

3. Methods for finding the jump resonance frequencies in terms of the system constants and the input amplitude are presented. The jump resonance frequencies obtained by these methods agree quite satisfactorily with the experimental values obtained by means of an electronic analog computer.

4. The closed-loop frequency-response characteristics of nonlinear feedback-control systems obtained by the present method agree quite favorably with the experimental results obtained by an electronic-analog-computer study. Since the equations for finding the closed-loop frequency response of feedback-control systems with on-off nonlinearities are not complicated, it may be concluded that this method may be used most advantageously for finding the closed-loop frequency-response characteristics of systems having on-off nonlinearities or high-degree saturation nonlinearities.

Appendix. Derivation of Equations for Finding Attenuation-Frequency and Phase-Frequency Characteristics of Nonlinear Second-Order Feedback-Control System with Saturation

Relation Between Output Velocity and Output Period

The block diagram of the system under consideration is shown in Fig. 3(A). The inverse Laplace transform of equation 4 gives the time solution of the system output velocity.

$$
\begin{aligned}
L^{-1}[V(s)] &= v(t) \\
&= \frac{KM_o}{d}\Bigg\{ t-\sum_{n=0}^{\infty}(-1)^n(t-d-an)\times \\
&\quad u(t-d-an)-\sum_{n=0}^{\infty}(-1)^n(t+d-a- \\
&\quad an)\times u(t+d-a-an)-T+T\sum_{n=0}^{\infty}\times \\
&\quad (-1)^n\times u(t-d-an)+T\sum_{n=0}^{\infty}(- \\
&\quad 1)^n\times u(t+d-a-an)+T\Bigg[\epsilon^{-t/T}- \\
&\quad \sum_{n=0}^{\infty}(-1)^n\epsilon^{-\frac{1}{T}(t-d-an)}\times u(t-d- \\
&\quad an)-\sum_{n=0}^{\infty}(-1)^n\epsilon^{-\frac{1}{T}(t+d-a-an)}\times \\
&\quad u(t+d-a-an)\Bigg]\Bigg\}+v(0)\epsilon^{-t/T} \qquad (33)
\end{aligned}
$$

where

$$
\left.\begin{aligned} u(t)&=0 \quad \text{when } t<0 \\ u(t)&=1 \quad \text{when } t>0 \end{aligned}\right\} \qquad (34)
$$

Equation 33 contains the initial velocity $v(0)$ term. To obtain $v(t)$ in a closed form and also to eliminate the $v(0)$ term in equation 33, let

$$\lambda a \leqslant t \leqslant (\lambda+1)a \qquad (35)$$

where λ is a sufficiently large number. Then $\epsilon^{-t/T}$ becomes negligibly small, with the result that the initial condition term in equation 33 may be eliminated.

$$v(0)\epsilon^{-t/T}\to 0 \qquad (36)$$

Arbitrarily choosing λ as an odd number, the value of $m(t)$ for $\lambda a<t<(\lambda+1)a$ becomes negative.

$$m[\lambda a<t<(\lambda+1)a]<0 \qquad (37)$$

Fig. 17 shows curve $m(t)$. Since curve

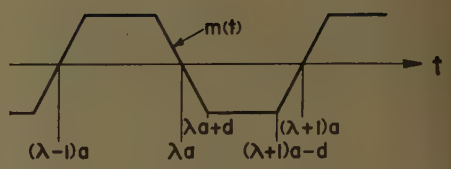


Fig. 17. Curve m(t)

$m(t)$ consists of straight-line segments as shown in Fig. 17, the time interval $\lambda a\leqslant t\leqslant(\lambda+1)a$ may be divided into the following three intervals corresponding to the three straight-line segments.

$$
\left.\begin{aligned} &\lambda a\leqslant t\leqslant \lambda a+d \\ &\lambda a+d\leqslant t\leqslant(\lambda+1)a-d \\ &(\lambda+1)a-d\leqslant t\leqslant(\lambda+1)a \end{aligned}\right\} \qquad (38)
$$

To simplify the derivation of equations, let

$$t-\lambda a\equiv\tau \qquad (39)$$

and substitute τ in equation 38, thus the three time intervals may be written as

$$
\left.\begin{aligned} &0\leqslant\tau\leqslant d \\ &d\leqslant\tau\leqslant a-d \\ &a-d\leqslant\tau\leqslant a \end{aligned}\right\} \qquad (40)
$$

The output velocity $v(\tau)$ for each time interval may be obtained as a function of τ, a, the half-period of the output, and d, as shown in the following.

For Time Interval $0\leqslant\tau\leqslant d$

By applying equation 34 to equation 33, the upper limits of the summation signs in equation 33 become $\lambda-1$ instead of ∞ for this time interval. Then equation 33 may be simplified as follows

$$
\begin{aligned}
v(0\leqslant\tau\leqslant d) &= v_1(\tau) \\
&= \frac{KM_o}{d}\Bigg\{(\tau+\lambda a)-\sum_{n=0}^{\lambda-1}(-1)^n(\tau+ \\
&\quad \lambda a-d-an)-\sum_{n=0}^{\lambda-1}(-1)^n(\tau+\lambda a+d- \\
&\quad a-an)-T+2T\sum_{n=0}^{\lambda-1}(-1)^n+ \\
&\quad T\Bigg[-\sum_{n=0}^{\lambda-1}(-1)^n\epsilon^{-\frac{1}{T}(\tau+\lambda a-d-an)}- \\
&\quad \sum_{n=0}^{\lambda-1}(-1)^n\epsilon^{-\frac{1}{T}(\tau+\lambda a+d-a-an)}\Bigg]\Bigg\} \\
&= \frac{KM_o}{d}\Bigg[-\tau+T-T\Bigg(\cosh\frac{d}{T}- \\
&\quad \sinh\frac{d}{T}\tanh\frac{a}{2T}\Bigg)\epsilon^{-\tau/T}\Bigg] \qquad (41)
\end{aligned}
$$

For Time Interval $d\leqslant\tau\leqslant a-d$

$$
\begin{aligned}
v(d\leqslant\tau\leqslant a-d) &= v_2(\tau) \\
&= KM_o\left(-1+\frac{T}{d}\epsilon^{\frac{a}{2T}-\frac{\tau}{T}}\times\frac{\sinh\frac{d}{T}}{\cosh\frac{a}{2T}}\right) \qquad (42)
\end{aligned}
$$

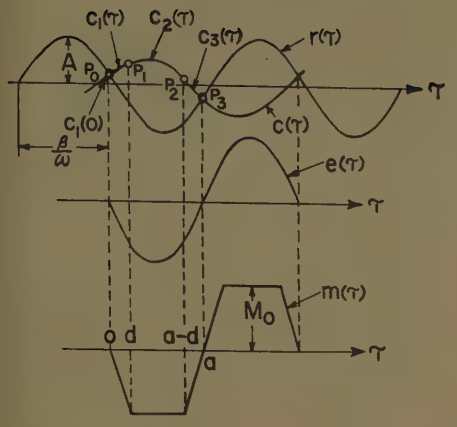


Fig. 18. Curves $c_1(\tau)$, $c_2(\tau)$, and $c_3(\tau)$ showing continuous and symmetric nature of $c(\tau)$

For Time Interval $a-d \leqslant \tau \leqslant a$

$$v(a-d \leqslant \tau \leqslant a) = v_3(\tau) = \frac{KM_0}{d}\left[(\tau-a)-T+T\times\left(\cosh\frac{d}{T}-\sinh\frac{d}{T}\tanh\frac{a}{2T}\right)\epsilon^{-\frac{\tau-a}{T}}\right] \quad (43)$$

Relation Between Output Position and Output Period

To find the expression for the output amplitude as a function of the output frequency, the output position is found as a function of τ, a, and d.

Since

$$c_i(\tau) = \int v_i(\tau)d\tau,\ i=1,\ 2,\ 3 \quad (44)$$

$c_1(\tau)$, $c_2(\tau)$, and $c_3(\tau)$ may be found by integrating equations 41 through 43 with respect to τ. Because of the continuous and symmetric nature of $c(\tau)$ as shown in Fig. 18, the following equations are obtained.

At point P_1 in Fig. 18

$$c_2(\tau \rightarrow d) = c_3(\tau \rightarrow d) \quad (45)$$

At point P_2

$$c_2(\tau \rightarrow a-d) = c_3(\tau \rightarrow a-d) \quad (46)$$

At points P_0 and P_3, $c_1(\tau)$ and $c_3(\tau)$ have the same magnitudes with the opposite signs. Therefore

$$c_1(\tau \rightarrow 0) = -c_3(\tau \rightarrow a) \quad (47)$$

Using equations 45 through 47, the system output may be found as follows

$$c(0 \leqslant \tau \leqslant d) = c_1(\tau) = \frac{KM_0T}{d}\left[\tau-T+T\times\left(\cosh\frac{d}{T}-\sinh\frac{d}{T}\tanh\frac{a}{2T}\right)\epsilon^{-\tau/T}\right]+\frac{KM_0}{2}(a-d)-\frac{KM_0\tau^2}{2d} \quad (48)$$

$$c(d \leqslant \tau \leqslant a-d) = c_2(\tau) = KM_0T\times\left(-\frac{\tau}{T}-\frac{T}{d}\epsilon^{\frac{a}{2T}-\frac{\tau}{T}}\times\frac{\sinh\frac{d}{T}}{\cosh\frac{a}{2T}}\right)+KM_0\left(\frac{a}{2}+T\right) \quad (49)$$

$$c(a-d \leqslant \tau \leqslant a) = c_3(\tau) = \frac{KM_0T}{d}\left[T-T\times\left(\cosh\frac{d}{T}-\sinh\frac{d}{T}\tanh\frac{a}{2T}\right)\epsilon^{\frac{a-\tau}{T}}-(\tau-a)+\frac{(\tau-a)^2}{2T}\right]-\frac{KM_0}{2}(a-d) \quad (50)$$

Substituting $\tau=0$ in equation 48, equation 10 may be obtained. Equations 48 through 50 show $c_1(\tau)$, $c_2(\tau)$, and $c_3(\tau)$ as functions of τ and a, the half-period of the output, and d. The amplitude of the output $c(\tau)$ may be found as the maximum of $c(0 \leqslant \tau \leqslant a)$ during the time interval $0 \leqslant \tau \leqslant a$ as shown in the following.

Relation Between Output Amplitude and Output Period

The maximum of $c(\tau)$ occurs when $v(\tau)$ becomes zero. If d/a is small and the velocity of the output becomes zero at some instant between the time interval d and $a-d$, the maximum of $c(\tau)$ occurs during the time interval $d \leqslant \tau \leqslant a-d$. Since τ_{max} was defined before as the time when the output position $c(\tau)$ becomes maximum, τ_{max} may be obtained by letting $v_2(\tau)=0$.

$$\tau_{max} = T\left(\ln\sinh\frac{d}{T}-\ln\frac{d}{T}-\ln\cosh\frac{a}{2T}+\frac{a}{2T}\right) \quad (51)$$

Let C_p be the output amplitude. Substituting equation 51 in equation 49, C_p or $c_2(\tau_{max})$ may be written as follows:

$$C_p = c_2(\tau_{max}) = KM_0T\times\left(\ln\cosh\frac{a}{2T}-\ln\sinh\frac{d}{T}+\ln\frac{d}{T}\right) \quad (52)$$

Equation 52 shows the output amplitude C_p as a function of a and d.

Relation Between Output Phase Lag and Output Period

By substituting $\tau=0$ in equation 48, equating $c(0)$ to $A \sin\beta$, and letting $\delta = 180°-\beta$, the following equation may be obtained:

$$\sin\delta = \frac{KM_0T}{Ad}\left[-T+T\times\left(\cosh\frac{d}{T}-\sinh\frac{d}{T}\tanh\frac{a}{2T}\right)+\frac{(a-d)d}{2T}\right] \quad (53)$$

From Fig. 4,

$$\theta_1 = 90°-\delta \quad (54)$$

$$\theta_2 = \tau_{max}\omega \quad (55)$$

Substituting equation 51 in equation 55, the phase lag ϕ which is the sum of θ_1 and θ_2 may be obtained as follows:

$$\phi = 180°\left[1-\frac{T}{a}\left(\ln\cosh\frac{a}{2T}+\ln\frac{d}{T}-\ln\sinh\frac{d}{T}\right)\right]-\delta \quad (56)$$

Substituting $a=\pi/\omega$ in equations 53 and 56, equations 7 and 8 may be obtained.

References

1. Nonlinear Control Systems, **N. Minorsky.** "Automatic and Manual Control." *Proceedings*, Cranfield Conference 1951, Butterworths Scientific Publications, London, England, 1952, pp. 309–18.

2. Some Saturation Phenomena in Servomechanisms with Emphasis on the Tachometer Stabilized System, **E. Levinson.** *AIEE Transactions*, vol. 72, pt. II, Mar. 1953, pp. 1–9.

3. A Generalized Method for Determining the Closed-Loop Frequency Response of Nonlinear Systems, **Luther T. Prince, Jr.** *Ibid.*, vol. 73, pt. II, Sept. 1954, pp. 217–24.

4. Automatic Feedback Control System Synthesis (book), **J. G. Truxal.** McGraw-Hill Book Company, Inc., New York, N. Y., 1955.

5. A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms, **Ralph J. Kochenburger.** *Ibid.*, vol. 69, pt. I, 1950, pp. 270–84.

6. Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements, **E. Calvin Johnson.** *Ibid.*, vol. 71, pt. II, July 1952, pp. 169–81.

7. Sub- and Super-Harmonic Oscillations of Nonlinear Control Systems, **K. Ogata.** *Ph.D. Thesis*, University of California, Berkeley, Calif., June 1956.

8. Modern Operational Mathematics in Engineering (book), **R. V. Churchill.** McGraw-Hill Book Company, Inc., 1944, p. 54.

Adaptive Servomechanisms

R. F. DRENICK
NONMEMBER AIEE

R. A. SHAHBENDER
NONMEMBER AIEE

Synopsis: A generalized approach is developed to aid in the synthesis of servomechanisms exhibiting adaptive behavior. The design methods described in this paper permit the synthesis of adaptive systems to follow polynomial signals up to any predetermined degree in the presence of Gaussian noise of known autocorrelation function. The basis of the method is the requirement that the servo loop adjust its parameters so as to minimize the steady-state root-mean-square error resulting from a polynomial input in the presence of noise. This condition specifies the functional relation between the servo parameters and the polynomial input. The physical realization of these functional relationships is treated in the paper and leads to the adaptation loops necessary to produce the adaptive behavior. As an example of the synthesis procedure, the design of a second-order adaptive system is presented. Analog computer results demonstrating that the adaptive system is superior in performance to linear systems of the same order are also included.

THE PURPOSE of this paper is to present an approach to the logical design of a class of servomechanisms which we have here designated as "adaptive." This term is borrowed from biology where it describes the ability of an organism to adjust itself to its environment. It is here used in the same sense. Specifically it refers to control systems which, roughly speaking, monitor their own performance and adjust some of their parameters in the direction of better performance.

It may be worth pointing out that such adjustments, whether in an organism or in a control system, need not be brought about by the self-monitoring principle suggested in the foregoing. It is often possible to monitor the environment directly and to adjust the control system accordingly. For example in the case of an aircraft auto-pilot, it is possible to determine the altitude of flight and to continuously adjust the auto-pilot for optimum performance. However, there exist situations in which the independent sensing of the environment is impractical or impossible. It is such cases that are dealt with in this paper.

The problem is more specifically the following. A control system is to be designed for a load with inertia and viscous damping, supplemented with a suitable equalizing network so that it can properly follow a signal embedded in noise. The signal, in Part II of the paper, is assumed to be a ramp function. The changing environment which can lead to the requirement for adaptation, is introduced by the fact that the slope of the ramp changes from time to time.

The manner in which the requirement for adaptation arises can be seen by the following argument. In choosing a suitable equalizing network, the designer is faced with a dilemna. He knows that it is necessary to include proportional plus integral action in the equalizing network in order to avoid steady-state tracking errors. In the case where the command signal is not contaminated with noise, the designer would undoubtedly choose such a mode of control. The parameters of the equalizing network are then chosen for acceptable transient performance.

When the incoming signal is contaminated with noise, the problem is more complicated. An equalizing network giving proportional control is more effective in smoothing the noise, than a network giving proportional plus integral action. However, the servo system with pure proportional control is not capable of properly tracking a ramp input. It is afflicted with a steady-state tracking error which offsets its superior noise smoothing ability.

A compromise is clearly desirable. The adaptive servomechanisms proposed in this paper strive, in a way, to attain this compromise. The simple proportional control is retained for the equalizing network. However, unlike a conventional servo, an additional loop, called the "adaptation loop" is added to the system. The purpose of this loop is to monitor the performance of the servo and to detect tracking errors in its output which are attributable to an inadequately followed ramp input, rather than to mere noise. When such an error is detected the adaptation loop changes the time constants of the servo and adjusts them to those values which represent the best compromise between noise-smoothing and tracking precision.

The control system which is obtained by this procedure performs gratifyingly well. When the ramps in its input are flat or totally absent, the system grows relatively sluggish and concentrates heavily on its noise smoothing task. As the ramps grow steeper the system progressively sacrifices sluggishness for tightness of response, and exchanges smoothness of operation for precision in tracking.

The design procedure developed in this paper pertains to a class of synthesis problems which is similar to that treated by Phillips.[1] However, the assumptions made in this paper and the results obtained are different. It is hoped, more particularly, that the procedure developed will be more easily carried out in practice and lead to mechanisms whose all-round performance is less controversial. This might make up for some of the conceptual shortcomings of the theory.

The paper is organized as follows: Part II contains the exposition in detail for the synthesis of the adaptive servo which has been used previously as the example for discussion. This is the servo with simple proportional control, a second-order servo, supplemented with an adaptation loop. The idea of the proper compromise with noise smoothing and precision tracking is first taken up. This is followed by a method for the design of the adaptation loop.

The case of the second-order servo is virtually the simplest conceivable case to which this procedure can be applied; yet it contains most of the relevant features of the general case. To facilitate the transition to the latter, an intermediate case is treated next, in Part III. This is the third-order servo, that is to say, a servomechanism in which the equalizing network incorporates proportional and integral control, both adjustable by an adaptation loop. The common features of the two examples point directly to the general case which is discussed briefly (Part IV).

In order to illustrate the usefulness of the procedure, a numerical example is discussed. Part V contains the analytical evaluation, as far as possible, of a second-order adaptive servo, along with two borderline competitors: conventional linear servos with simple proportional control, one synthesized with emphasis on noise-smoothing, the other on precision tracking. Part VI contains a discussion of some analog computer runs on the same

Paper 57-388, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer General Meeting, Montreal, Que., Canada, June 24-28, 1957. Manuscript submitted July 5, 1956; made available for printing April 24, 1957.

R. F. DRENICK is with Bell Telephone Laboratories, Inc., New York, N. Y., and R. A. SHAHBENDER is with the Radio Corporation of America, Camden, N. J.

The authors wish to acknowledge the interested assistance of the staff of Electronics Associates Computation Center, Princeton, N. J., in setting up the computer simulation and in obtaining the experimental results.

This work was supported in part by the U. S. Air Force under Contract *AF 33(600)28007*.

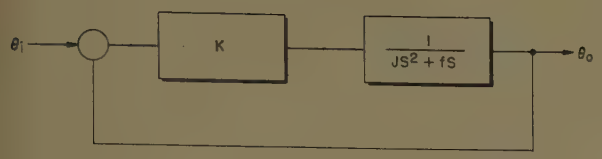


Fig. 1 (left). Proportional control system

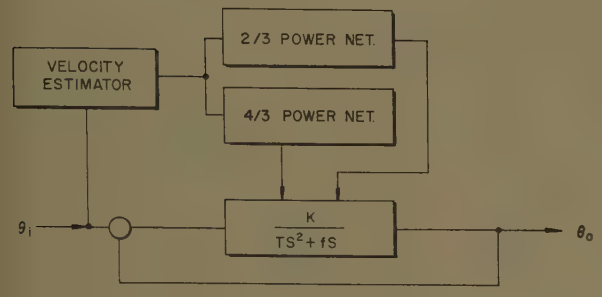


Fig. 2 (left). Second-order adaptive servo system

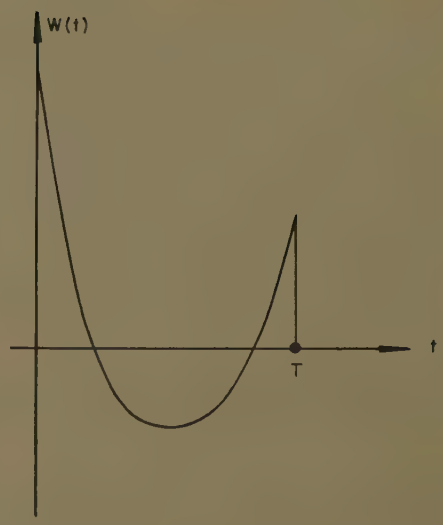


Fig. 3 (right). Optimum weighting function for velocity estimator

examples. Both discussions show that the adaptive servo combines the advantages of both competitors to a gratifying degree.

II. Second-Order Systems

Preliminary Remarks

In this section an outline will be given of the approach to the design of an adaptation loop for a positional servo with proportional control (Fig. 1), that is, a servo whose linear version is described by the second-order transfer function

$$\frac{\theta_o(s)}{\theta_i(s)}=\frac{\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2} \qquad (1)$$

$$\omega_n^2=K/J$$

$$\zeta=f/2\sqrt{KJ}$$

where

$\theta_i(t)$ = input to servo system
$\theta_o(t)$ = output of servo system
ω_n = angular frequency of undamped oscillation
ζ = damping ratio
J = motor and load inertia
f = viscous damping coefficient
K = motor amplifier gain (proportional gain or spring constant)
s = Laplace variable

The input to the servo, that is, to the linear as well as to the adaptive versions, consists of a ramp signal

$$S(t)=\begin{cases}A_r t & (t\geq 0)\\ 0 & (t<0)\end{cases} \qquad (2)$$

and Gaussian noise of known autocorrelation function $R_N(\tau)$. In fact, since the design procedure to be outlined in the following is feasible regardless of the form of the autocorrelation function, it will be sufficient to assume for it its simplest form

$$R_N(\tau)=(1/2)\sigma_N^2\delta(\tau) \qquad (3)$$

corresponding to white noise.

The performance of the servos, linear as well as adaptive, will be judged by the square error in the steady state. This error has two components: the mean-square error due to the noise (ϵ_N^2) and the square of the servo lag in following the signal(ϵ_s^2). The total square error (ϵ_T^2) is accordingly

$$\epsilon_T^2=\epsilon_s^2+\epsilon_N^2 \qquad (4)$$

and will serve in what follows as performance and optimization criterion.[2]

Parameter Variation of Second-Order Systems

For the linear second-order proportional control system defined in the foregoing, the total error in equation 4 in following a ramp input is

$$\epsilon_T^2=\frac{4\zeta^2}{\omega_n^2}A_r^2+\frac{\omega_n}{8\zeta}\sigma_N^2$$

For fixed ζ, the minimum value of ϵ_T^2 is attained when:

$$\omega_n=\frac{4\zeta}{(\sigma_N^2)^{1/3}}A_r^{2/3}$$

and if, in addition, the inertia is fixed, this requires

$$K\alpha A_r^{4/3},\ f\alpha A_r^{2/3} \qquad (5)$$

Therefore, to have optimum operation, the slope of the ramp signal would have to be determined, a task which is impossible, strictly speaking, because the noise overlies the signal. Instead of an exact determination, however, one can resort to a statistical estimation procedure. This will be relegated to the adaptation loop.

Optimum Adaptation Loop for Second-Order Systems

The adaptation loop for the second-order system consists of a first-derivative (or velocity) estimator, followed by networks that give the required powers (equation 5) of the estimator output as shown in Fig. 2. Thus the adaptation loop design problem reduces to the choice of a suitable velocity estimator. The weighting function $W(t)$ of the optimum estimator, which may be determined by the method of Zadeh and Ragazzini,[3] leads, as shown in Appendix I, to the following result:

$$W(t)=\frac{36}{T^2}-\frac{192}{T^3}t+\frac{180}{T^4}t^2 \quad 0\leq t\leq T$$
$$=0 \quad \text{elsewhere.} \qquad (6)$$

Fig. 3 is a plot of the foregoing function. Since $W(t)$ is discontinuous at $t=T$, it is not realizable in terms of lumped elements; and it becomes necessary, as will be discussed later, to determine a realizable approximation to this weighting function.

The behavior of the optimum estimator in equation 6 is evidently governed by the choice of T, which can be effected by the following line of reasoning. Since the output of the estimator attains its steady-state value after an interval of T seconds from the application of a ramp input, it is desirable to make T as short as possible. On the other hand, the mean-square noise σ_E^2 at the output of the estimator network is

$$\sigma_E^2=\frac{96}{T^3}\sigma_N^2$$

which implies as large a T as possible.

From these considerations it is possible to fix a suitable compromise value for T. As shown in the appendix, such a value is

$$T=\left[\frac{96\alpha_1^2\sigma_N^2}{A_1^2}\right]^{1/3}$$

where

α_1 = minimum acceptable signal-to-noise ratio for the estimator
A_1 = smallest value of input velocity to be detected

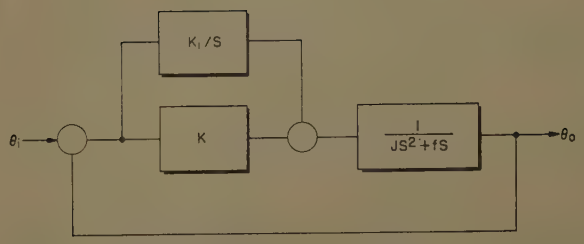


Fig. 4 (above). Proportional-plus-integral control system

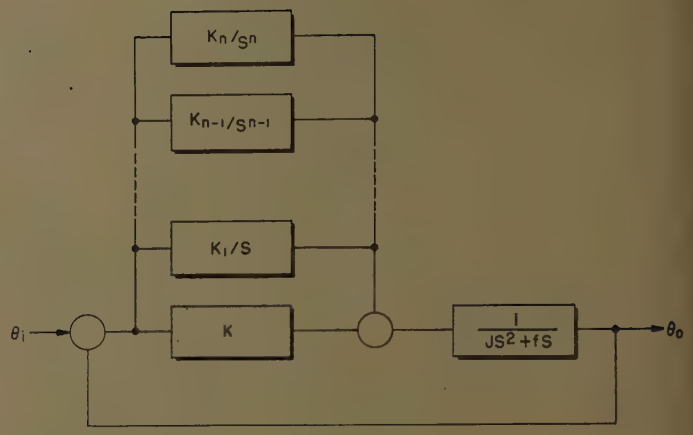


Fig. 5 (right). Proportional-plus-multiple integral control system

Differential Equation of Adaptive Servo

The adaptive servo just synthesized obeys a differential equation of the form

$$\dddot{\theta}_o + C_1\theta_E^{2/3}\ddot{\theta}_o + C_2\theta_E^{4/3}\dot{\theta}_o = C_3\theta_E^{4/3}\theta_i$$

where θ_E has been written for the output of the estimator network

$$\theta_E(t) = \int_{-\infty}^{t} W(\tau)\theta_i(t-\tau)d\tau$$

This is an interesting type of equation: It is linear in the sense that the dependent variable θ_o and its derivative enter linearly only; but it is nonlinear in that it violates the super-position principle.

Its stability is now an only partially answered question. It can be proved stable under a very large class of inputs, including step and ramp functions; but it is not known, for instance, whether stability is assured under sinusoidal inputs regardless of frequency.

Physically Realizable Adaptation Loop for Second-Order System

The weighting function of equation 6 of the optimum velocity estimator is not realizable in terms of lumped elements. To determine a weighting function $W_a(t)$ that approximates the optimum function and that is realizable in terms of lumped elements, apply Foster's t-method. Let

$$W_a(t) = \sum_{k=1}^{3} B_k e^{-\alpha_k t}$$

where the B_k's and α_k's are constants to be determined so that $W_a(t)$ approximates $W(t)$. The following set of conditions may be used to determine these constants.

1. $W_a(t)$ and $W(t)$ have equal moments.

$$\int_0^\infty t^i W_a(t)dt = \int_0^T t^i W(t)dt \quad i = 0, 1, \text{and } 2$$

2. $W_a(t)$ has a tail area that is 1% of the total area:

$$\int_T^\infty W_a(t)dt = 0.01\int_0^\infty W_a(t)dt$$

3. $W_a(0) = W(0)$

4. $W_a(t)$ has minimum mean-square noise output when the input consists of the noise signal $N(t)$.

It may be shown from the foregoing conditions that it is possible to determine the values of the constants associated with the weighting function and thus realize the estimator in terms of lumped elements.

An example of this procedure is given in Section V.

III. Third-Order Systems

Preliminary Remarks

The preceding discussion concerning the proportional control servo can readily be extended to the next case in order of complication: the positional servo with proportional-plus-integral control (Fig. 4). This is described by a third-order transfer function of the form:

$$\frac{\theta_o(s)}{\theta_i(s)} = \frac{(\omega_n^2 + 2\zeta\omega_n\gamma)s + \gamma\omega_n^2}{(s^2+2\zeta\omega_n s+\omega_n^2)(s+\gamma)}$$

$$2\zeta\omega_n + \gamma = f/J$$

$$2\zeta\omega_n\gamma + \omega_n^2 = K/J$$

$$\gamma\omega_n^2 = K_1/J$$

where

K = proportional gain
K_1 = integrator gain
γ = real pole of the transfer function

The input is now assumed to be a signal of the form

$$S(t) = \begin{cases} (1/2)A_q t^2 & (t \geq 0) \\ 0 & (t < 0) \end{cases} \qquad (7)$$

contaminated with white Gaussian noise, as described above. The performance criterion is again the total steady state, squared error ϵ_T^2.

Parameter Variation

For the third-order system, the total error in following the quadratic input of equation 7 is given by the expression:

$$\epsilon_T^2 = \left[\frac{2\zeta\omega_n+\gamma}{\omega_n^2\gamma}\right]^2 A_q^2 + \frac{\omega_n}{8\zeta}\left[1+\frac{4\zeta\gamma(\omega_n+\zeta\gamma)}{\omega_n^2+2\zeta\omega_n\gamma+\gamma^2}\right]\sigma_N^2$$

The above expression includes three parameters (ζ, ω_n, and γ) that may, in principle, be varied to produce adaptive behavior. The step response of the third-order system consists of an exponential term plus a damped sinusoidal term. The damping of the sinusodial term is governed by ζ, which, having been chosen for good transient recovery, may be assumed fixed.

The parameter γ specifies the exponential term in the step response of the system. As shown in the literature,[1] the choice of

$$\gamma = 2\zeta\omega_n$$

leads to a system with good performance characteristics. For that value of γ we have:

$$\epsilon_T^2 = \frac{4}{\omega_n^4}A_q^2 + \frac{\omega_n}{8\zeta}\sigma_N^2 + \frac{1+2\zeta^2}{1+8\zeta^2}\zeta\omega_n\sigma_N^2$$

The minimum value of ϵ_T^2 is then obtained when

$$\omega_n = 2\left[\frac{4\zeta(1+8\zeta^2)}{(1+\zeta+8\zeta^2+2\zeta^3)\sigma_N^2}\right]^{1/5} A_q^{2/5}$$

Since J is assumed constant, and σ_N^2 is constant, the above expression may be written as:

$$\omega_n = CA_q^{2/5}$$

where C is a constant. Further, we have:

$$\gamma = 2\zeta CA_q^{2/5}$$

From the relations between J, f, K, K_1 and ζ, ω_n, γ as given previously—again assuming unalterable inertia—one can determine the dependence on the acceleration A_q of the proportional gain K, the integrator gain K_1, and the viscous damping f of the servo system:

$$f\alpha A_q^{2/5}, \qquad K\alpha A_q^{4/5}, \qquad K_1\alpha A_q^{6/5}$$

Since the servo parameters are propor-

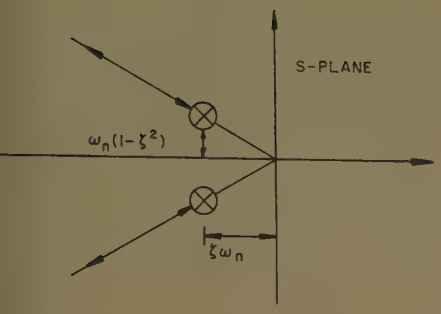


Fig. 6. Pole motion for second-order adaptive system

tional to these powers of A_q, the adaptation loop must consist of: (1) an acceleration estimator, and (2) networks to produce powers of the estimator output.

Optimum Adaptation Loop

The adaptation loop for the third-order system consists of an acceleration estimator in tandem with networks that give the required powers of the estimator output. The ideal weighting function of the estimator is given by:[3]

$$W(t)=\frac{480}{T^3}-\frac{5{,}400}{T^4}\,t+\frac{12{,}960}{T^5}\,t^2-\frac{8{,}400}{T^6}\,t^3 \qquad 0\le t\le T$$

$$=0 \quad \text{elsewhere}$$

which must, in practice, be again approximated by a physically realizable function.

The mean-square noise output of the estimator is given by:

$$\sigma_E^2=\frac{12.96\times10^3}{T^5}\,\sigma_N^2$$

As in the case of the second-order system, the estimator must have a minimum signal-to-noise ratio α_2 for a smallest detectable acceleration magnitude A_2. The estimator time constant is then given by:

$$T=\left[\frac{12.96(10^3\alpha_2^2\sigma_N^2}{A_2^2}\right]^{1/5}$$

IV. Generalization to Higher-Order Servomechanisms

The generalization of the notions exhibited above leads to a positional servo with proportional-plus-multiple-integral control (Fig. 5) with the transfer function

$$\frac{\theta_i(s)}{\theta_o(s)}=\frac{Ks^n+K_1s^{n-1}+\ldots+K_n}{Js^{n+2}+fs^{n+1}+Ks^n+K_1s^{n-1}+\ldots+K_n}$$

where

K = proportional gain
K_i = gain of ith integrator ($i=1, 2, \ldots n$)

An nth order servo system, as defined above, will exhibit no steady-state errors, other than noise errors, for polynomial inputs of degree less than, or equal to, $(n-2)$; constant steady-state errors for polynomial inputs of degree $(n-1)$; and steady-state errors that increase with time for inputs of degree n or higher. For inputs of degree n the steady-state error grows as t, and for inputs of degree $(n+1)$ as t^2— and so on.

The effect of the adaptation loops which were proposed above for second- and third-order systems can be illustrated by the motion of the poles, which is induced in the transfer functions of the linear system by the adaptive behavior. This idea can then be carried over to the design of higher-order systems.

A second-order servo system is characterized by a closed-loop transfer function having a single pair of complex conjugate poles in the s-plane. The adaptation loop operates so as to increase ω_n from its unadapted value in the presence of ramp inputs but maintaining ζ constant. Thus the adaptation loop operates so as to move the poles into the interior of the left-half of the complex plane along the two straight lines shown in Fig. 6. Similarly, in the case of a third-order system the closed-loop transfer function is characterized by a single pair of complex conjugate poles and a real pole. The adaptation loop operates so as to move the poles into the interior of the left half of the complex plane along the paths shown in Fig. 7.

A fourth-order system may be treated in the same manner—except for two cases that must be considered.

1. The closed-loop transfer function has two pairs of complex conjugate poles with corresponding damping ratios ζ_1, and ζ_2 and angular frequencies ω_1 and ω_2. The adaptation loop operates so as to move the poles in a manner similar to that shown in Fig. 6.

2. The closed-loop transfer function has a single pair of complex conjugate poles and two real poles, γ_1 and γ_2. For this case:

$$\gamma_1=C_1\omega_n$$

$$\gamma_2=C_2\omega_n$$

C_1 and C_2 are constants. The adaptation loop operates so as to move the poles in a manner similar to that shown in Fig. 7.

This can be generalized to systems of any order. However, it is of purely academic interest to consider, of the type under consideration, any systems higher than the fourth order.

For a second-order system the estimator consists of a network that determines the first derivative of the input. For a third-order system the estimator determines the second derivative, and so on. The procedure employed in determining the time constant associated with the adaptation loop for the second- and third-order systems may obviously be extended to systems of higher order.

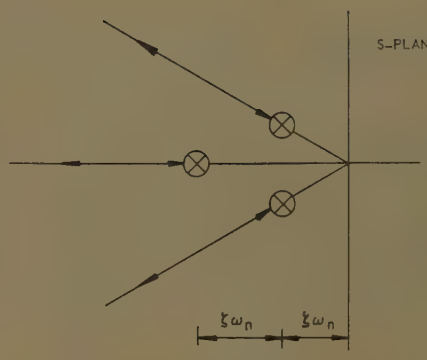


Fig. 7. Pole motion for third-order adaptive system

V. Design of a Specific Second-Order Adaptive System

For the second-order adaptive system the time constant associated with the velocity estimator is given by:

$$T=\left[\frac{96\alpha_1^3\sigma_N^2}{A_1^2}\right]^{1/3}$$

The relation between ω_n and A_r is given by:

$$\omega_n=\frac{4\zeta}{(\sigma_N^2)^{1/3}}\,A_r^{2/3}$$

Assuming the following values for the quantities given in the above expressions:

$\sigma_N^2=0.636$ square foot per second
$A_1=20$ feet per second
$\alpha_1=0.5$
$\zeta=0.5$

we obtain:

$T=0.534$ second
$\omega_n=2.33\,A_r^{2/3}$

The weighting function of the optimum estimator is completely determined by the value of T given above. The approximate weighting function may now be determined from the conditions given in Section II. One finds:

$$W_a(t)=250e^{-t/.089}-158e^{-t/.178}+15.8e^{-t/.356}$$

Let $\theta_E(t)$ denote the output of the estimator for an input $\theta_i(t)$. The adaptive-system damping and proportional gain are given by:

$$f/J=2.33\theta_E^{2/3}$$

$$K/J=5.43\theta_E^{4/3}$$

where J is assumed constant.

Table I

Input, Feet/Second	Adaptive System: Sampling	No. of Samples	T, Seconds	τ, Seconds	RMS Error, Feet	Linear Tracker: Sampling	No. of Samples	T, Seconds	τ, Seconds	RMS Error, Feet
Noise	Uniform	243	15.1		1.38	Uniform	200	5		2.65
Noise+20	Uniform	122	4.2	3.75	2.56	Uniform	121	4.2	3.75	2.69
Noise+50	Random	200	5	5	2.62	Random	200	5	5	2.64
Noise+100	Random	200	2	5	2.99	Random	200	2	5	3.04

Table III

Input Code Letter	Description of Input: Increasing Ramp, Ft/Sec	Decreasing Ramp, Ft/Sec
A	Noise+20	20
B	Noise+20	80
C	Noise+80	20
D	Noise+80	80

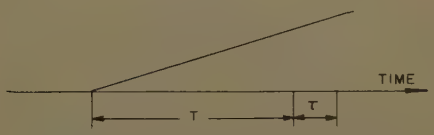


Fig. 8. Ramp input adaptive system

The evaluation of the adaptive system may be carried a step further by comparing its performance with the performance of the optimum filter that will reproduce a ramp input when contaminated with white noise. The weighting function for such a filter may be determined by the method of Zadeh and Ragazzini,[3] and is:

$$W(t)=\frac{4}{T}-\frac{6}{T^2}\,t \qquad 0\leq t\leq T$$

$$=0 \text{ elsewhere.}$$

The mean square noise output σ_o^2 of such a filter, when subjected to the same noise input as the adaptive system is

$$\sigma_o^2=\frac{1.272}{T}$$

Inspection of the above expression shows that values of T greater than about 1/2 second are necessary in order that the optimum filter have a root-mean-square noise output comparable to the adaptive system. For such values of T, the transient recovery period of the optimum filter is considerably longer than that of the adaptive system. In addition, since the weighting function of the optimum filter is discontinuous at time T, it is not realizable in terms of lumped elements. Thus an approximate weighting function must be determined which is then realized in terms of lumped elements.

VI. Analog Computer Simulation of Adaptive Servo System

In order to evaluate the performance of the adaptive servo, the example of the preceding section was simulated, along with two linear systems, on an Electronics Associates analog computer. The linear systems were chosen as the two extremes: the one to represent a good smoothing system with correspondingly large time constants, and the other a good tracking system with small time constants. All three were subjected to the same inputs, and records of the output errors were obtained.

More particularly, the two linear systems were

1. Good tracker

$\omega_n=2\pi\times 8$ radians per second, $\zeta=0.5$

2. Good smoother

$\omega_n=2\pi\times 1$ radians per second, $\zeta=0.5$

Because of equipment limitations it was possible only to simulate the adaptive system, together with one linear system, in each sequence of tests.

The simulated systems were subjected to the following inputs: (1) noise plus ramps, as shown in Fig. 8; and (2) noise plus an increasing ramp followed by a decreasing ramp, as shown in Fig. 9.

The noise signal was derived from a noise generator with a spectral density that was essentially constant (adjusted to 0.636 square foot per second) up to a frequency of 60 cycles per second.

In accordance with the objective of the simulation (to provide a quantitative comparison between adaptive and linear systems), the error signal, that is, the

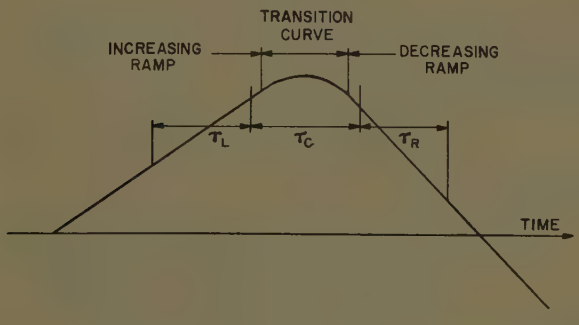


Fig. 9 (left). Combination ramp input to adaptive system

Table IV

Input	Adaptive System Errors, Feet: ϵ_L*	ϵ_C*	ϵ_R*	Linear Tracker Errors, Feet: ϵ_L	ϵ_C	ϵ_R
A	2.66	2.28	2.34	2.75	2.67	3.07
B	2.41	2.78	3.37	2.41	2.70	2.92
C	3.63	2.98	2.47	2.84	2.63	2.19
D	3.34	3.06	3.52	3.22	2.84	2.98

* ϵ_L, ϵ_C, and ϵ_R are the rms errors in tracking combination of ramp inputs, as shown in Fig. 9, over intervals τ_L, τ_C, and τ_R, respectively.

Table II

Input, Feet/Second	Adaptive System: Sampling	No. of Samples	T, Seconds	τ, Seconds	RMS Error, Feet	Linear Smoother: Sampling	No. of Samples	T, Seconds	τ, Seconds	RMS Error, Feet
Noise	Uniform	200	9.25		1.46	Uniform	200	9.25		1.03
Noise+20	Random	200	5.63	5	2.57	Uniform	81	5.63	5	3.87
Noise+50	Random	200	1.5	5	3.18	Uniform	81	1.5	5	8.91
Noise+100	Random	100	3	2.5	3.63	Uniform	49	2.5	3	17.00

Table V

Input	Adaptive System Errors, Feet: ϵ_L	ϵ_C	ϵ_R	Linear Smoother Errors, Feet: ϵ_L	ϵ_C	ϵ_R
A	2.08	1.88	2.92	3.71	3.49	3.81
B	2.35	2.42	3.15	3.57	8.23	11.30
C	3.22	1.72	1.85	8.55	3.68	3.58
D	3.62	3.36	3.91	14.3	10.03	13.9

difference between the servo output and the noiseless ramp input, or combination of ramp inputs, was recorded in each case. Since the error signal is a random function, it was sampled either in uniform intervals of τ seconds or, in some cases, at random intervals; this fact is noted in the tabulated results. Time T, which was allowed for transients to die out, is also noted.

Table I, which contains the comparison between the adaptive servo and the good linear tracker, shows not only that the former smooths noise in the absence of a signal better than the latter by a factor of two, but also shows that it performs equally well in tracking ramps through noise. Table II exhibits corresponding results for the adaptive servo compared with the good linear smoother. The smoothing properties of the two are here shown to be about the same in the case of noise only; in tracking a ramp through noise, however, the adaptive servo is quite superior.

Similar results for combinations of ramp inputs are contained in Tables III, IV, and V.

In order to evaluate qualitatively the transient behavior of the adaptive system, responses to step inputs were simulated. At the outset it was feared that, since the design procedure was aimed chiefly at good steady-state behavior, the transient response might develop to be poor. The opposite was found to be true. The recovery time of the servo appeared to be substantially shorter than that of even the good linear tracker and exhibited a damping ratio equivalent to about $\zeta = 0.5$.

Appendix

Weighting Function of Optimum Velocity Estimator

Following the procedure derived by Zadeh and Ragazzini,[3] let $W(t)$ be the weighting function of the estimator

$W(t)=0$ for $t<0$ and $t>T$,

and let

$$\mu_i=\int_0^\infty dt\, t^i W(t)=\int_0^T dt\, t^i W(t)$$

be the ith moment of $W(t)$. Let the estimator be subjected to a signal consisting of a polynomial $P(t)$ of, at most, the second degree and a random noise with autocorrelation.

$$R_N(\tau)=(1/2)\sigma_N{}^2\delta(\tau)$$

and spectral density

$$G_N(f)=\sigma_N{}^2$$

For a velocity estimator

$$\mu_0=0 \qquad \mu_1=-1 \qquad \mu_2=0$$

The weighting function is given by

$$W(t)=A_0+A_1t+A_2t^2 \qquad 0\le t\le T$$
$$=0 \text{ elsewhere}$$

where A_0, A_1, and A_2 are constants determined from the moments of $W(t)$. Thus,

$$A_0T+\frac{A_1}{2}T^2+\frac{A_2}{3}T^3=\mu_0$$

$$\frac{A_0}{2}T^2+\frac{A_1}{3}T^3+\frac{A_2}{4}T^4=\mu_1$$

$$\frac{A_0}{3}T^3+\frac{A_1}{4}T^4+\frac{A_2}{5}T^5=\mu_2$$

Solving the above equations, we obtain

$$A_0=\frac{36}{T^2}, \quad A_1=\frac{192}{T^3}, \quad A_2=\frac{180}{T^4}$$

Hence,

$$W(t)=\frac{36}{T^2}-\frac{192}{T^3}t+\frac{180}{T^4}t^2 \qquad 0\le t\le T$$
$$=0 \text{ elsewhere}$$

Choice of the Estimator Time Constant

If the estimator is subjected to an input consisting of a noise $N(t)$ and a polynomial of the form

$$P(t)=A_0+A_1t$$

The steady-state output will consist of a noise with root-mean-square value σ_E superposed on the derivative of $P(t)$, i.e., A_1.

Let α_1 be the signal-to-noise ratio at the estimator output. Then,

$$\alpha_1{}^2=\frac{A_1{}^2}{\sigma_E{}^2}=\frac{T^3}{96}\frac{A_1{}^2}{\sigma_N{}^2}$$

or

$$T=\left[\frac{96\alpha^2\sigma_N{}^2}{A_1{}^2}\right]^{1/3}$$

Thus if α_1 is the minimum acceptable signal-to-noise ratio for the estimator, then A_1 is the smallest value of velocity input that can be detected. The estimator time constant T is then given by the above expression.

References

1. Theory of Servomechanisms, H. M. James, N. B. Nichols, R. S. Phillips. *Radiation Laboratory Series*, Massachusetts Institute of Technology, Cambridge, Mass., vol. 25.

2. An Adaptive Servo System, A. H. Benner, R. F. Drenick. *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 4, 1955, p. 8.

3. An Extension of Wiener's Theory of Prediction, L. A. Zadeh, J. R. Ragazzini. *Journal of Applied Physics*, New York, N. Y., vol. 21, July 1950.

Discussion

Ian McCausland (University of Toronto, Toronto, Ont., Canada): The authors define an adaptive control system as one which monitors its own performance and adjusts some of its own parameters in the direction of better performance. However, any feedback control system may be said to monitor its own output or outputs and to adjust something in the system in such a way as to decrease the discrepancy between what the performance is and what it ought to be. To call one system adaptive because it varies parameters, and to exclude another system from this category because it only varies signals, seems to be an undue restriction of the meaning of the word "adaptive." Is the variation of parameters the essential feature of an adaptive system? If so, an ON-OFF servomechanism is adaptive since it varies the loop gain from zero to a definite positive or negative value in accordance with fixed rules. If we include an ON-OFF servomechanism in the adaptive category can we exclude the proportional-control system, which provides a continuously-variable adjustment of the gain between the power-supply and the load?

The authors state that the term "adaptive" has been borrowed from biology. However, in the science of biology this term describes a number of different types of process. For example, Ashby[1] proposes the definition that "a form of behavior is adaptive if it maintains the essential variables within physiological limits"; and many of the examples given by him to illustrate adaptive behaviour are not essentially different from the conventional feedback control system. The meaning of the term "adaptation" in the theatrical field seems to be closer to what the authors have in mind, that is, the alteration of a piece of dramatic writing in order to suit the taste of a different audience or to accommodate it to different circumstances.

Since interest in the type of servomechanism described by the authors is likely to increase in the future, it appears to be necessary at this stage to choose technical terms which are as precise as possible. Since the term "adaptive" is used so widely in biology it seems desirable to qualify it in some way, so that it will not be taken to describe the entire class of feedback control systems.

Reference

1. Design for a Brain (book), W. Ross Ashby. John Wiley & Sons, Inc., New York, N. Y., 1954, chap. 5.

R. F. Drenick and R. A. Shahbender: Mr. McCausland raises a point of terminology. He agrees that control systems of the type discussed by us belong to a class of such systems and that they differ from the conventional type. However, he feels that the term "adaptive" is overly comprehensive and therefore a poor choice.

Points of terminology are always difficult to settle but the first thing to do in any case is to try to define the class of systems to

which a term is to be attached. As far as we know, this has not been done. We shall therefore make an attempt at it. It will of necessity be a rather broad and unspecific definition, partly because the state of our knowledge about these systems, and partly because the nature of this discussion, preclude going into much detail. Nevertheless, it will become quite clear that the class as we hope to define it will include the servo mechanisms which we have discussed, along with many others; that it does not include the ON-OFF system or the proportional-control system mentioned in his comments. Finally, we hope to make plausible our choice of the term "adaptive" for the class of systems.

Broadly speaking, we expect the servo mechanisms of this class to be designed for operation in a slowly changing environment, as opposed to the conventional types which are intended for a fixed environment. To be more specific, we must define what we mean by "environment." We use the term here collectively for all the data, given or assumed, which must be taken into account by the system designer, that is:

1. The nature of the signal for which the system is to be designed.
2. The nature of the noise (if any) against which it should discriminate.
3. The set of parameters (if any) of the system which are given at the outset and which cannot be negotiated by the designer.

It is correct to say, we believe, that a conventional servomechanism is intended for operation in a fixed environment, or in any case, is designed with such an environment in mind. That is to say, the nature of the signal and of the noise (their spectral densities, say) and of the nonnegotiable parameters (some of the gains and time constants) are fixed or, if necessary, are artificially frozen at some reasonable values. The control system is then so designed that it will operate in this environment in some optimal, or nearly optimal, fashion.

We propose that, in order for a control system to belong to the class which we have called adaptive, two conditions should be fulfilled, namely:

1. The environment should be allowed to change slowly (when compared with the frequencies of signal and noise), the system should be able to detect changes and vary its (negotiable) parameters accordingly.
2. The system should have a single input, comprising signal and possibly noise, and it should be designed to derive all indications of environmental change either from that input, or from its own response to it, or both.

We omit discussion of these specifications or of their motivation. It is probably worth pointing out that the class is rather broad, that it includes the type of system we dealt with in our paper, and many others, but that it does not include ON-OFF servos nor conventional linear servos (except possibly as a pointless border line case).

We can now comment briefly on the matter of terminology. To our knowledge, the following terms have been attached to control systems essentially of the class defined here: "input-controlled," "sympathetic," "learning," "self-adjusting," "self-designing," and "adaptive." The last, which was our choice, is used in biological fields in the sense of "modification of an animal or plant (or its parts or organs) fitting it more perfectly for existence under the condition of its environment," according to Webster's Collegiate Dictionary. We felt, and still feel, that this describes very well the control systems under discussion.

Quasi-Optimization of Relay Servos by Use of Discontinuous Damping

W. L. HARRIS, JR. STUDENT MEMBER AIEE
C. McDONALD STUDENT MEMBER AIEE
G. J. THALER MEMBER AIEE

A NUMBER of articles[1–6] have been written which deal with the optimum response of a relay servo to a step input, with some mention of a ramp input. If optimum response is defined as that response which regains position correspondence in minimum time and without overshoot, then it has been shown that an ideal relay servo with second order motor-load combination may be optimized. The optimized relay servo has one period of acceleration with maximum torque, an instantaneous torque reversal (ideal relay) at an appropriate point, and one period of deceleration with maximum torque. The system attains zero error and zero error-rate simultaneously, the torque is removed, and the system is stationary at zero error. For the second order motor-load, with a maximum torque specification, this represents the fastest possible response. For higher order motor-load units, more switching operations[5] are required.

In order to realize such an optimized servo, an anticipating device is required to reverse the torque at the proper instant. Phase plane analysis shows that the locus of points on the phase plane at which this reversal is required is precisely that deceleration trajectory which passes through the origin of the phase plane. This is shown in Fig. 1. Physical realization of this switching locus requires a nonlinear anticipator. Satisfactory operation of the servo depends on the accuracy with which the nonlinear device produces the desired switching locus, and also depends on the properties of the relaying device, which has been postulated as ideal.

A quasi-optimum relay servo is hereby defined as one which has all of the characteristics of an optimum relay servo except that of minimum response time. In practice, an ideal relay does not exist. Assuming that the only defect in the relay is the existence of a dead zone, a quasi-optimum relay servo may be obtained by the application of a damping or braking torque when the relay opens. If the damping is properly controlled, the system comes to rest at or near zero error, and it will be shown that this can be accomplished for a wide range of step inputs. The resulting system is deadbeat, has one accelerating period under full torque, and one decelerating period with variable torque as a result of the natural damping plus the discontinuously applied damping.

Nomenclature

C = coulomb friction torque in pound-feet
D = added viscous damping torque
$E = \theta_R - \theta_c$ = position error in radians
E_{ss} = steady-state position error in radians
f = coefficient of viscous friction in pound-feet per radian per second
I_a = motor armature current in amperes
J = moment of inertia in slug-feet squared
K_t = motor torque constant in pound-feet per ampere
K_v = motor back emf constant in volts per radian per second
L_a = armature inductance in henries
R_a = armature resistance in ohms
R_B = dynamic braking resistance in ohms
T = motor shaft torque in pound-feet
V = voltage applied through relay in volts
θ_c = output shaft position in radians
θ_R = reference shaft position in radians

Paper 57-605, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Great Lakes District Meeting, Des Moines, Iowa, April 15–17, 1957. Manuscript submitted January 8, 1957; made available for printing May 17, 1957.

W. L. HARRIS, JR., C. McDONALD, and G. J. THALER are with the United States Naval Postgraduate School, Monterey, Calif.

The opinions contained herein are the private ones of the authors and are not to be construed as official or reflecting the views of the Navy Department or the Naval Service at large.

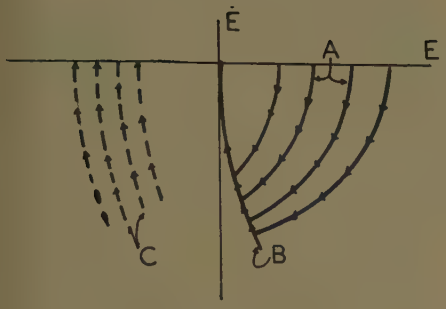


Fig. 1. Phase trajectories for an optimum relay servo

A—Acceleration trajectories
B—Deceleration trajectories
C—Deceleration trajectory and locus of relay reversals

Phase Plane Analysis of Discontinuous Damping

For the purpose of introducing the basic theory, it is assumed that the relay servo consists of an ideal error detector, pure torque motor, and pure inertia load. The basic differential equation during acceleration periods is

$$J\ddot{E} = +T \tag{1}$$

If the relay has dead zone, when the relay opens the differential equation is

$$J\ddot{E} = 0 \tag{2}$$

The phase trajectory corresponding to these conditions is sketched in Fig. 2(A). Note that the system traverses the dead zone at constant velocity. If the motor-load unit contains some viscous damping, then equations 1 and 2 are altered and the effect on the phase trajectory is shown by dotted curve of Fig. 2(A). Note that the damping causes some deceleration in the dead zone. If some damping is added to the output at the instant the relay opens, then the output is decelerated as it crosses the dead zone. Furthermore, if the added damping is viscous damping, then equation 2 becomes

$$J\ddot{E} = -D = -f\dot{E} \tag{3}$$

Such discontinuous introduction of damping does not affect equation 1, and therefore permits maximum acceleration when the motor is driving. The resultant phase trajectory is shown on Fig. 2(B). Note that the deceleration trajectory in the dead zone is a straight line. This is a convenient and fortuitous circumstance. If an appreciable amount of coulomb friction is present, this deceleration trajectory is curved.

Addition of a linear differentiating device, either in the error channel or in feedback, rotates the dead zone boundaries. Such a condition is illustrated in Fig. 2(C). Note that this provides anticipatory switching according to a linear law. From the phase trajectory of Fig. 2(C), the effect is obviously stabilizing. If the magnitude of the coefficient of the discontinuous viscous damping is considered in determining the rotation to be produced by the differentiating device, the phase trajectory in the dead zone can be made precisely parallel to the dead zone boundary, and the trajectory falls exactly on this boundary, as shown in Fig. 2(D). The system is then deadbeat for all magnitudes of step input. The response time depends on the magnitude of the viscous damping, as well as the magnitude of the step input, and increasing the damping coefficient permits reduction in response time.

The deadbeat response shown in Fig. 2(D) has a steady state error E_{ss} which is equal to one half of the total dead zone width. However, since the trajectory does not enter the dead zone but merely follows its boundary, the width of the dead zone may be decreased until the dead zone width is infinitesimal and the steady state error is negligible, i.e., $E_{ss} \cong 0$. This has been illustrated on Fig. 2(E). This, then, represents a quasi-optimized condition.

Note that if the viscous damping is made sufficiently large, the dead zone approaches a vertical position and the response time may be even shorter for the quasi-optimized system then for the so-called optimum relay servo.

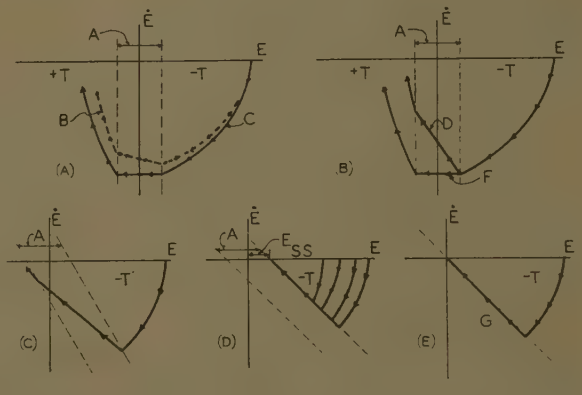


Fig. 2. Phase trajectories describing the effect of discontinuous damping

A—Trajectories of basic relay servo
B—Effect of discontinuous damping
C—Rotation of dead zone by differentiator
D—Matching the slopes of trajectory and dead zone boundary
E—Reducing width of dead zone to optimize response

Discontinuous Damping by Dynamic Braking of a D-C Shunt Motor

The theory presented in the preceding section is readily applied to a relay servo using a d-c shunt motor. Discontinuous damping is obtained by connecting a dynamic braking resistor across the armature terminals of the motor, as shown in Fig. 3. Note that the only extraneous effect during acceleration intervals is to draw additional current through the relay contacts, but this does not affect the dynamic performance of the motor.

Assuming that the armature inductance and the coulomb friction are negligible and that the relay is ideal, it may be shown that the phase plane equations are

$$\frac{d\dot{E}}{dE} = -\frac{f}{J} - \frac{K_T K_v}{JR} - \frac{K_T V}{JR\dot{E}} \tag{4}$$

for the acceleration period, and

$$\frac{d\dot{E}}{dE} = -\frac{f}{J} - \frac{K_T K_v}{J(R_a R_B)} \tag{5}$$

for the deceleration in the dead zone. (The symbols are defined in the nomenclature section, and the equations are derived in Appendix I.) Equations 4 and 5 define the isoclines on the phase plane. Equation 4 is typical of torque saturated operation and needs no comment, but equation 5 shows that the trajectory in the dead zone is a straight line, and that the slope of this line may be adjusted by varying R_B. Since these equations show that the d-c shunt motor with dynamic

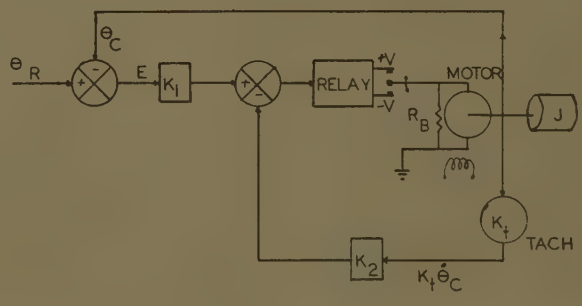


Fig. 3. Block diagram of a shunt motor relay servo with dynamic braking for discontinuous damping

Fig. 4. The 1-hp motor load unit with associated control and measurement devices

braking has precisely the phase plane characteristics required by the theory, actual performance will vary from the theoretical only if the assumptions are not valid. Experimental verification of the theory is, therefore, necessary.

Characteristics of D-C Servomotors, Coulomb Friction Effects

It is obvious that relays in general have some time lag in operation, motor armatures possess inductance,[7] and motor-load combinations have some coulomb friction. Thus, it is known that the three basic assumptions are not completely valid. However, if these effects are negligible, or if they may be minimized by proper design, the resulting system performance can be satisfactory The coulomb friction effect is readily checked prior to assembling the servo. The motor, dynamic braking resistor, gear train, tachometer, and error detector are assembled as they are to be used in the servo, thus, including all friction effects. The motor is run at saturation speed, then the supply voltage is removed from the armature and the motor coasts to standstill. Velocity and position measurements are recorded with a two channel Brush recorder, and replotted to obtain the phase trajectory.

Such retardation tests were run on three sizes of d-c shunt servomotors. A 1-hp (horsepower) motor-load unit is shown in Fig. 4. Retardation tests were run on this unit, using several values of R_B. A phase trajectory is plotted on Fig. 5. Note that this is nearly a straight line; the only perceptible coulomb effect is the slight hook as the velocity approaches zero. For this motor it would appear that coulomb friction is negligible. Similar tests were run on a 1/8-hp motor and a 1/125-hp motor, and typical curves have been added to Fig. 5. Note that these added curves are each for nearly maximum dynamic braking and represent the best trajectories obtainable. Note also, that the co-ordinates have been adjusted by proportion to facilitate comparison of the trajectories' curvatures. From Fig. 5 it is seen that the largest motor shows the least coulomb friction effect, its trajectory being an almost perfect straight line. Also, the smallest motor shows the greatest coulomb friction effect, its trajectory being considerably curved. It can be shown mathematically that this variation in curvature is largely due to the C/J ratio; see Appendix II.

For all of the cases shown in Fig. 5, the coulomb friction effect would prevent the phase trajectory from following the dead-zone boundary exactly. However, practical relays have a pull-in point which requires more voltage than is needed to hold the relay closed, so that the practical relay drops out at a lower voltage then the pull-in voltage. The phase portrait therefore requires' additional boundary lines for the dead zone, as shown in Fig. 6(A), and for slight curvature of the deceleration trajectory the system is completely deadbeat, but with slightly increased static error, as also shown on Fig. 6(A). For larger amounts of coulomb friction, the forward driving relay may close several times producing a jerky motion. The phase trajectory is shown in Fig. 6(B). The net result is some decrease in the static accuracy, since a finite dead zone is required to obtain deadbeat operation. The actual width of this dead zone depends on the specific system.

Effect of Relay Time Lag and Armature Inductance as Shown by Step Function Tests on the Relay Servo

The 1-hp motor-load unit was selected for the initial investigation because of the negligible coulomb friction effect. It was incorporated in a relay servo, the circuit being shown in Fig. 7. Note that control of the dead zone was easy because the amplifiers used were taken from an analog computer and provided a very flexible arrangement. The phase trajectories obtained in response to a step input are shown in Fig. 8. The departure from theoretical predictions is obvious, but is readily explained and may be corrected to a large extent. One major discrepancy is the lack of a sharp change in the slope of the trajectory at the drop-out line. However, a fairly sharp change occurs farther along the trajectory, in the dead zone. This suggests a time lag in relay operation, and auxiliary tests verified this. The importance of minimizing relay lag thus becomes obvious. A second discrepancy in the trajectories of Fig. 8 lies in the fact that the trajectory does not become a straight line immediately after the relay opens, but continues curving before it straightens out. This is caused by armature inductance. Fig. 9 shows a curve of armature current versus time. The relay closes at A, and the current rises rapidly (but not instantaneously) to a maximum, then decays along B to a non-zero steady state at C. When the relay drops out, the current must first decrease to zero, then reverse, going through a maximum along the curve CD. Note that a finite time is required for the current to decrease to zero, reverse, and rise to a maximum in the reverse direction. During this time interval, the dynamic braking effect is not as large as predicted, and the phase trajectory curves out into the dead zone. For smaller disturbances, the relay drops out earlier, i.e., nearer B on Fig. 9, and the effect of the armature inductance has a greater duration because it takes longer to

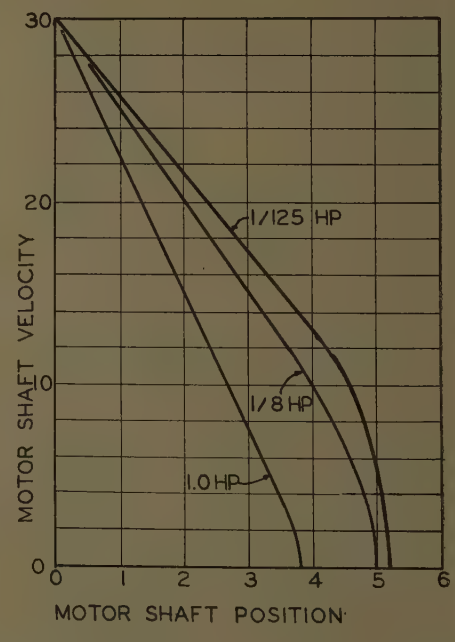


Fig. 5. Effect of coulomb friction on the phase trajectories of dynamically braked d-c shunt motors

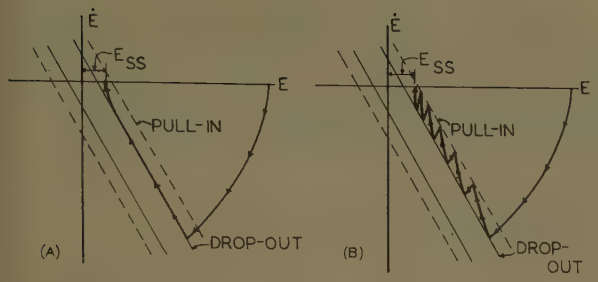


Fig. 6. Effect of coulomb friction on a relay servo when the relay has unequal pull-in and drop-out voltages

A—Effect of a small amount of coulomb friction
B—Effect of a large amount of coulomb friction

reduce the initial armature current to zero.

It is immediately obvious that the relay time lag can be decreased by using a better relay, and the delay due to armature inductance can be reduced by selecting a motor such that the electrical time constant of the armature, L_a/R, is small. Another factor to be noted is that both effects may be compressed on the phase plane by increasing the motor-load gear ratio. The wide dead zone of Fig. 8 is largely due to the fact that the gear ratio was chosen to emphasize the defects, and was only 13 to 1, so that an input step of 25 degrees was corrected in less than one revolution of the motor shaft.

Step and Ramp Function Tests on the Relay Servo Using a 1/125-Hp Motor

In order to permit use of a smaller and faster relay, the 1/125-hp motor replaced the larger motor. This motor also had a much more favorable electrical time constant for its armature, and finally the gear ratio was raised to 500 to 1. The phase trajectories for the response of this system to a step input are shown in Fig. 10. It is readily seen that these trajectories more nearly approximate those predicted by theory, and the width of the dead zone has been greatly reduced. Still further improvements can be made by means of motor selection, relay selection, and gear ratio adjustment.

This system was also tested with a ramp input. Due to the use of tachometer feedback, a velocity lag error is introduced. Substitution of an error differentiator eliminated the velocity lag error. The use of a ramp input always causes a limit cycle, which is characterized by opening and closing of the forward driving relay only; the reverse relay never operates. It is thus seen that the output during ramp operation is essentially confined to the region between the pull-in and the drop-out points. The frequency of the limit cycle increases with increased input velocities.

Comparison of a Quasi-Optimized Servo with an Optimum Servo Using the Same Motor-Load Unit

Fig. 11(A) shows an experimentally obtained phase trajectory for the 1/125-hp quasi-optimized servo, and also an optimum phase trajectory for the same motor-load unit obtained from experimental acceleration and decleration trajectories. There is an appreciable difference between these trajectories in the deceleration region, so Fig. 11(B) shows the transient response curves obtained from these trajectories. It is readily seen that the responses are identical during acceleration, nearly identical for part of the deceleration period, but the quasi-optimized system approches correspondence more slowly when the error has been reduced to a small value.

Conclusions

A theory has been developed for using discontinuous damping to obtain a relay servo which has a deadbeat response to a step input. This theory has been applied to the case of a relay servo using a d-c shunt motor, in which the discontinuous damping is obtained by connecting a dynamic braking resistor across the motor armature. It has been shown that static accuracy is adversely affected by relay time lag, by the electrical time constant of the armature, and by coulomb friction, but it has also been shown that these effects can be controlled and an acceptable system obtained. It is obvious that the quasi-optimized system is simpler to design, construct, and adjust than a comparable system using optimum switching, since the latter requires a special nonlinear anticipator for switching. It has also been shown that the transient response of the quasi-optimized system is not too unfavorable as compared with the optimum response. Finally, it might be noted that the dynamic braking resistor tends to lock the output and prevent wander in the dead zone when the system is at rest.

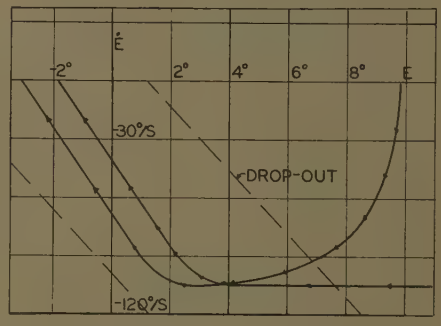


Fig. 8. Phase trajectories of a quasi-optimized relay servo using a 1-hp shunt motor (step response)

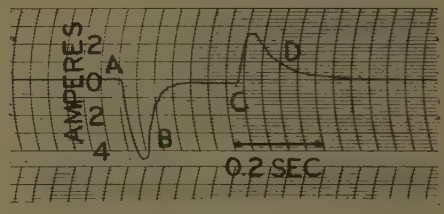


Fig. 9. Armature current variation in a dynamically braked relay servo

AB—Initial inrush
BC—Decrease to steady state as motor accelerates
CD—Reverse current due to dynamic braking

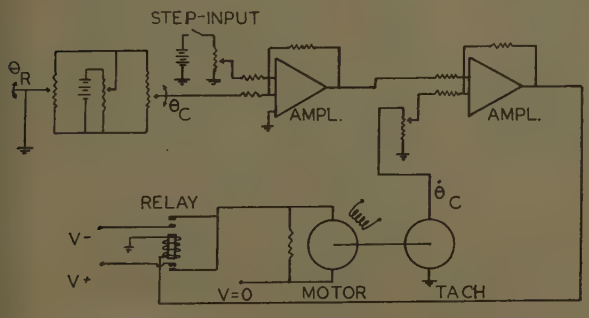


Fig. 7 (left). Circuit diagram of quasi-optimized relay servo using 1/125-hp d-c shunt motor

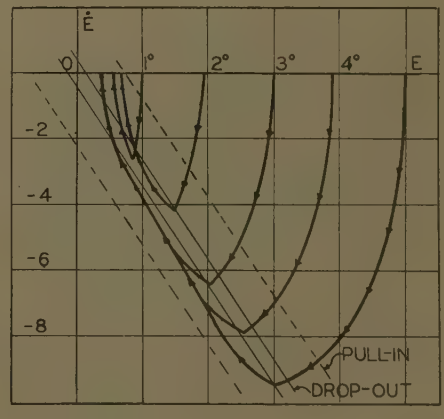


Fig. 10 (right). Phase trajectories showing the step response of a quasi-optimized relay servo using a 1/125-hp shunt motor

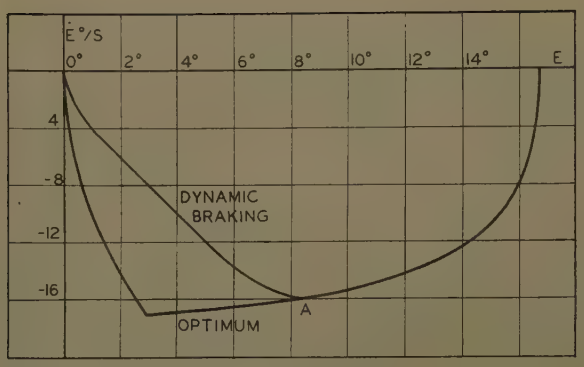


Fig. 11(A) (left). Comparison of phase trajectories for optimum and quasi-optimum operation of a relay servo

A—Point of relay drop-out for the quasi-optimum system

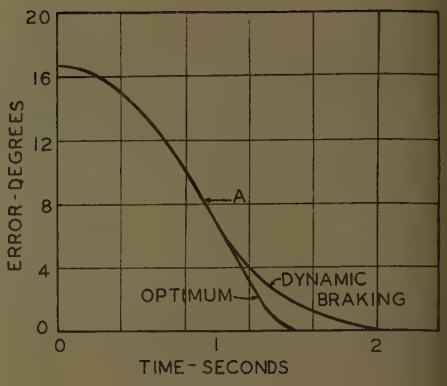


Fig. 11(B) (right). Comparison of time response of an optimum and a quasi-optimum relay servo

A—Point of relay drop-out for the quasi-optimum system

Appendix I

Derivation of Equations 4 and 5

Assumptions: C/J is negligible; $4R_a$ is negligible; the input is a step of position, θ_R. Applying Kirchhoff's voltage law,

$$V=K_v\dot{\theta}_c+I_aR_a \tag{6}$$

and for dynamic torque equilibrium,

$$K_tI_a=J\ddot{\theta}_c+f\dot{\theta}_c \tag{7}$$

but

$$E=\theta_R-\dot{\theta}_c \tag{8}$$

and for $t>0$, θ_R is constant, so

$$\dot{E}=-\dot{\theta}_c \quad (t,0)$$
$$\ddot{E}=-\ddot{\theta}_c \quad (t,0) \tag{9}$$

From which

$$V=-K_v\dot{E}+I_aR_a \tag{10}$$

$$K_tI_a=-J\ddot{E}-f\dot{E} \tag{11}$$

which may be combined to give the following:

$$\ddot{E}=-\frac{K_tV}{JR_a}-\frac{K_tK_v\dot{E}}{JR_a}-\frac{f}{J}\dot{E} \tag{12}$$

Dividing both sides by $dE/dt=\dot{E}$,

$$\frac{\frac{d^2F}{dt^2}}{\frac{dE}{dt}}=\frac{\frac{d\dot{E}}{dt}}{\frac{dE}{dt}}=\frac{d\dot{E}}{dE}=\frac{-K_tV}{JR_a\dot{E}}-\frac{K_tK_v}{JR_a}-\frac{f}{J} \tag{4}$$

which is the equation for the isoclines during the accelerating period.

For the deceleration in the dead zone due to dynamic braking, the following equations apply:

$$-K_v\dot{E}+I_aR_a+I_aR_B=0 \tag{13}$$

$$K_tI_a=-J\ddot{E}-f\dot{E} \tag{14}$$

These combine to give:

$$E=\frac{-K_tK_v\dot{E}}{J(R_a+R_B)}-\frac{f\dot{E}}{J} \tag{15}$$

Again dividing both sides by $dE/dt=\dot{E}$

$$\frac{\frac{d\dot{E}}{dt}}{\frac{dE}{dt}}=\frac{d\dot{E}}{dE}=\frac{-K_tK_v}{J(R_a+R_B)}-\frac{f}{J} \tag{5}$$

which is the equation for the isoclines in the dead zone.

Appendix II

Effect of Coulomb Friction on the Deceleration Trajectory

In this case, equation 13 is valid, but equation 14 must include a coulomb friction term, and becomes:

$$K_tI_a+J\ddot{E}+f\dot{E}-C=0 \tag{16}$$

Combining equations 13 and 16,

$$E=-\frac{K_tK_v}{J(R_a+R_B)}\dot{E}-\frac{f}{J}\dot{E}+\frac{C}{J} \tag{17}$$

Dividing both sides by $dE/dt=\dot{E}$,

$$\frac{\frac{d\dot{E}}{dt}}{\frac{dE}{dt}}=\frac{d\dot{E}}{dE}=-\frac{K_tK_v}{J(R_a+R_B)}-\frac{f}{J}+\frac{C}{J\dot{E}} \tag{18}$$

This equation for the isoclines shows that the phase trajectory curves as $\dot{E}$ varies, providing $C/J\neq 0$.

References

1. A Phase-Plane Approach to the Compensation of Saturating Servomechanisms, Arthur M. Hopkin. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 631–39.

2. Nonlinear Techniques for Improving Servo Performance, D. McDonald. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 6, 1950, pp. 400–21.

3. A Study of Some Nearly Optimum Servomechanisms, T. M. Stout. *Report no. 10*, Electrical Engineering Department, University of Washington, Seattle, Wash., 1952.

4. Optimization of Nonlinear Control Systems by Means of Nonlinear Feedbacks, R. S. Neiswander, R. H. MacNeal. *AIEE Transactions*, vol. 72, pt. II, Sept. 1953, pp. 262–72.

5. An Investigation of the Switching Criteria for Higher Order Contactor Servomechanisms, Irving Bogner, Louis F. Kazda. *Ibid.*, vol. 73, pt. II, July 1954, pp. 118–27.

6. Optimum Switching Criteria for Higher Order Contactor Servo With Interrupted Circuits, S. S. L. Chang. *Ibid.*, vol. 74, pt. II, Nov. 1955, pp. 273–76.

7. Transfer Function and Parameter Evaluation for D-C Servomotors, G. J. Thaler, W. A. Stein. *Ibid.* (Jan. 1956 section), pp. 410–17.

8. Discontinuous Damping of Relay Servomechanisms, W. L. Harris, Jr. *M.S. Thesis*, U. S. Naval Postgraduate School, Monterey, Calif., 1956.

Discussion

Frederick W. Nesline, Jr. (Massachusetts Institute of Technology, Cambridge, Mass.): The paper's contribution is the adjustment of the dead-zone trajectories to be parallel to the dead-zone boundaries so as to achieve step function or disturbance response without overshoot. This response is then compared to that of an optimum relay servomechanism. Such a comparison is not a good one because the constraints necessary to derive the optimum system used for comparison are violated by the quasi-optimum system. It is more logical to compare the quasi-optimum system with the optimum system that has the same constraints, i.e., discontinuous damping must be allowed in both cases. This true optimum does have a vertical phase trajectory along the $\dot{E}$ axis and exhibits the property of having the fastest possible time response to certain inputs for this set of constraints. Using the true optimum as a basis for comparison, the paradoxical statement that "the response time may even be shorter for the quasi-optimized system than for the so-called optimum relay servo" may be avoided.

Silver-Zinc Batteries as Source-Primary Electric Power for Pilotless Aircraft

LYAL T. ABRAHAMSON
ASSOCIATE MEMBER AIEE

THE design of a pilotless aircraft electric system must be approached in a far different manner than for a piloted airplane. The pilotless aircraft has nobody to substitute for inoperative motors, relays, or servo systems. Each component in the aircraft system must operate properly or the result very likely is an aborted mission. Also, at the speeds and accelerations encountered in pilotless aircraft, a tremendous amount of energy is expended to get each pound of equipment to its destination, thus an ever-increasing premium is placed on component weight. To achieve the conflicting requirements of high reliability and light weight with equipments either available or within reach of reasonable development time and efforts, the system depicted by Fig. 1 was designed and developed.

Note that a single-string system is used to keep weight to a minimum. This scheme depends on exceptionally high reliability in each component right down to the last nut, screw, and solder joint. It also follows that fuses are not required where single power sources supply systems which are essential to successful mission completion.

General System Description

A battery, power contactor, d-c motor, and variable displacement pump furnish hydraulic power for movement of the control surfaces. The battery is made up of a single string of series-connected secondary silver-zinc cells. Circuit redundancy is provided by multiterminal cells and parallel wires in the connecting cables. The d-c compound wound motor is directly connected to external connectors through which power is provided for manufacturing tests and prelaunch checkouts. Just prior to launching, the battery is connected to the motor by actuating the power contactor.

Paper 57-491, recommended by the AIEE Chemical Industry and Air Transportation Committees and approved by the AIEE Technical Operations Department for presentation at the AIEE East Central and Middle Eastern District Meeting and Air Transportation Conference, Dayton, Ohio, May 7–9, 1957. Manuscript submitted February 4, 1957; made available for printing May 23, 1957.

LYAL T. ABRAHAMSON is with Boeing Airplane Company, Seattle, Wash.

Power for the guidance and telemeter systems is furnished by three electrically separate battery supplies which are the +6.75-volt filament battery, the +28-volt inverter battery, and the multivoltage high-voltage battery. In addition to the 400-cycle-per-second 115/200-volt inverter, the heat-exchanger fan motor, small relays, and motors are supplied by the +28-volt battery. All three batteries are made up of single strings of series-connected secondary silver-zinc cells. Circuit redundancy is limited to the use of parallel pins in miniature connectors. As in the hydraulic system, aircraft checkouts are performed on externally supplied power. Prior to launching, the batteries are connected to their respective loads by the power changeover switch and power contactors.

Except in cases where power source isolation is required, all regulated power is provided by centralized voltage regulators. The (+250 ± 1%)-volt and (+150 ± 1%)-volt regulators are supplied by the high-voltage battery, and the (−250 ± 1%)-volt regulator derives its primary power from the inverter.

Silver-Zinc Batteries

Construction and Chemistry

The silver-zinc electrochemical system, although known for many years to be a highly concentrated source of electric power, was impractical until recently because of the very limited storage life. Discovery of an ion-selective separator material and subsequent improvements to it have made possible cells of the type utilized in the power system shown in Fig. 1.

In the discharged state the positive plates are made up of porous silver which is mechanically strong enough to maintain its original shape even when removed from the cell. The cathodes consist of zinc hydroxide powder contained in paper sacks. The electrolyte is a 40% solution of potassium hydroxide and water. The separator material surrounds the silver plates and prevents migration of silver ions to the cathode. The Bakelite *C-11* cases are designed to allow leak-free operation in any position, but in storage for extended periods should be upright.

A simplified over-all cell reaction is given by the following equation:[1]

$$AgO + Zn + H_2O \underset{\text{Discharged}}{\overset{\text{Charged}}{\rightleftharpoons}} Zn(OH)_2 + Ag$$

The reaction actually involves an intermediate state of the positive plate. As the charging begins the silver is converted to silver oxide (Ag_2O), as evidenced by a cell voltage of approximately 1.65 volts. When some of the silver oxide is converted to silver peroxide (AgO), the cell voltage increases to approximately 1.90 volts. The voltage rises above 2.05 volts when the cell is fully charged and oxygen evolution begins at the positive plate.

General Characteristics

In addition to providing a high concentration of potential energy, the silver-zinc cells exhibit a flatter discharge characteristic than most other electrochemical couples. As shown in Figs. 2 and 3, voltage tolerances of ±6% or less can be maintained until the capacity of the cell is nearly completely exhausted. In fact, moderate load increases such as those shown in Fig. 3 can be supplied during the last half of the battery discharge at a voltage equal to that near the beginning of the discharge. In contrast, the nickel-cadmium and lead-acid cells' output voltages continue to droop throughout their discharge. At discharge rates which expend the stored energy in approximately one hour, the silver-zinc cells provide from 30 to more than 50 watt-hours per pound, whereas the nickel-cadmium and lead-acid cells provide less than 10 watt-hours per pound.

The major limitation to the use of silver-zinc cells has been their comparatively short charged-storage life. Through continued research, the charged-storage life of secondary silver-zinc cells has been extended to more than a year. In applications where the batteries are not required to be kept in a constant state of readiness, battery life can be extended by refrigerated storage in the discharged condition.

Two by-products of the silver-zinc cell are of concern in their application. First, during charged storage, hydrogen is liberated at the cathode as a result of oxidation of the zinc by water. The amount of hydrogen liberated is not great, but caution must be taken in handling of batteries after long periods of charged storage. Second, near the end of high-rate discharges, the electrolyte boils quite violently and electrolyte is expelled from the cells. Thus installation provisions must be made to protect sensitive components from electrolyte spillage.

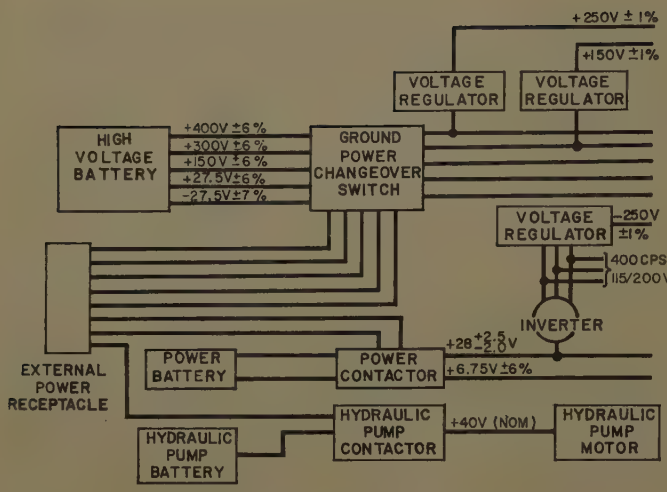


Fig. 1. Power-system block diagram

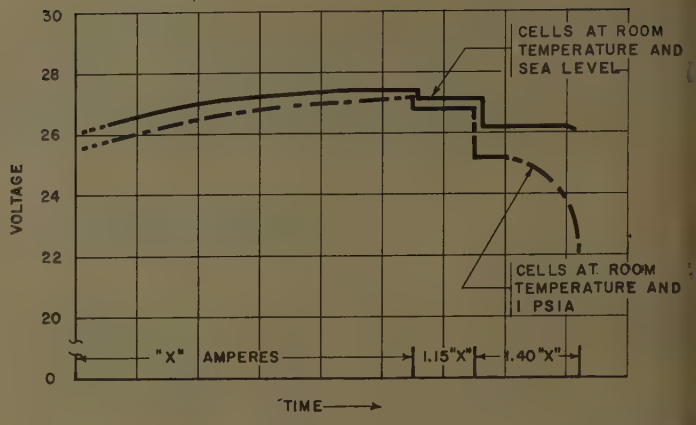


Fig. 3. Effect of altitude on cell voltage and capacity

Handling Techniques

Constant current charging consistent with the manufacturer's recommendations is used with all batteries. Whenever the cells are discharged for storage, rates taking from 2 to 4 hours are used to minimize heating of the cells.

Since the storage life of silver-zinc cells is limited, a constant search is made for handling techniques which will make the fullest use of their capabilities. The cells are purchased dry, in which condition they have been stored with no ill effects for over 3 years. When the need for a battery arises, the electrolyte is added and the formation cycle consisting of a low-rate charge and exhaustion discharge is performed. From this time until expenditure of the battery, deterioration continues. By storing the cells at approximately 20 to 30 F (degrees Fahrenheit) in the discharge condition, the useful life can be extended to three or four times what it is when the cells are stored at room temperature in the charged condition. Therefore the cells are maintained in the charged condition only when the need for a charged battery is present.

Another practical means found to increase the life of charged silver-zinc cells is replacing of the air in the battery cases with nitrogen. To investigate the benefits of nitrogen storage, 25 small cells (*HR-05*) were charged and placed in a sealed case containing nitrogen. After 3 months storage at 80°F, output efficiencies of from 67% to 81% were obtained at a 500-milliampere discharge rate. Recharge of the cells showed the input capacity was 81% of what it had been preceding the test. For comparison, 25 control cells were stored in air beside the nitrogen-filled case. At the end of the 3-month period, all the control cells failed to give any useful output as well as failing to respond to the recharge. Further tests using larger cells showed that the storage life could be increased more than 50% through the use of nitrogen-filled cases. Exclusion of oxygen from the cell apparently is the reason for the increase in cell life, since another series of tests using argon as a storage atmosphere gave similar results.

It has also been found that if the electrolyte is maintained at a high level in the cells, the life is increased markedly. However, electrolyte leakage then becomes a major problem and use of this scheme is impractical.

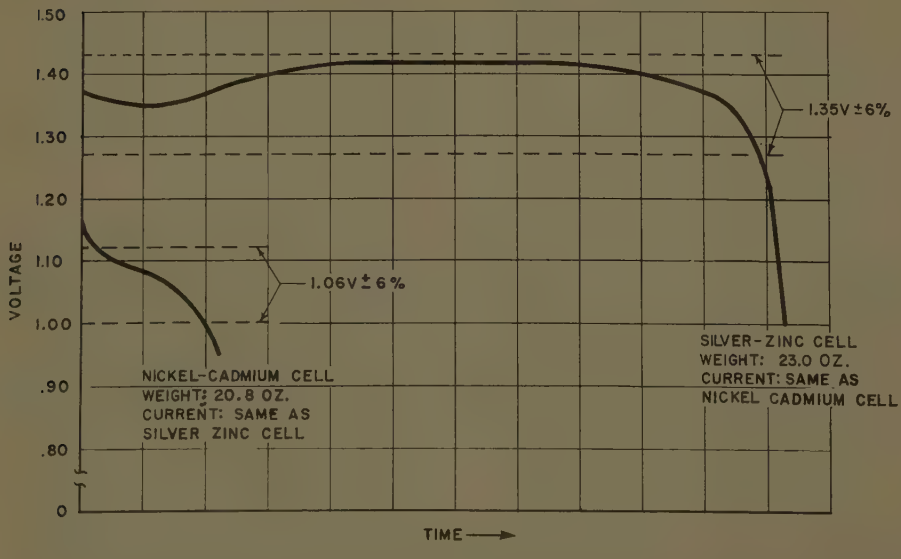


Fig. 2. Comparison of typical silver-zinc and nickle-cadmium cell discharge characteristics

Cell Selection

Two major considerations must be given in selection of a battery for any application. First, the battery must have sufficient ampere-hour capacity; and second, its voltage regulation must be consistent with the needs of the load.

The peak hydraulic power demanded is approximately three times that of the steady-state load. To keep the pump speed within the range of 6,200 to 10,000 rpm, the combined motor and battery weight was found to be lightest using a battery with excess ampere-hour capacity. In contrast, the loads of the other three batteries are more constant, and cells chosen for proper voltage regulation also have approximately the right ampere-hour capacity.

Application Problems

As with all components considered for installation in any aircraft, evaluation tests which include vibration, shock, temperature, humidity, altitude, and life tests are performed on the batteries. Two environmental conditions for which application modifications were found advisable are temperature and altitude.

During development of the battery for the +28-volt and 6.75-volt systems, protection of the electronic packages in ad-

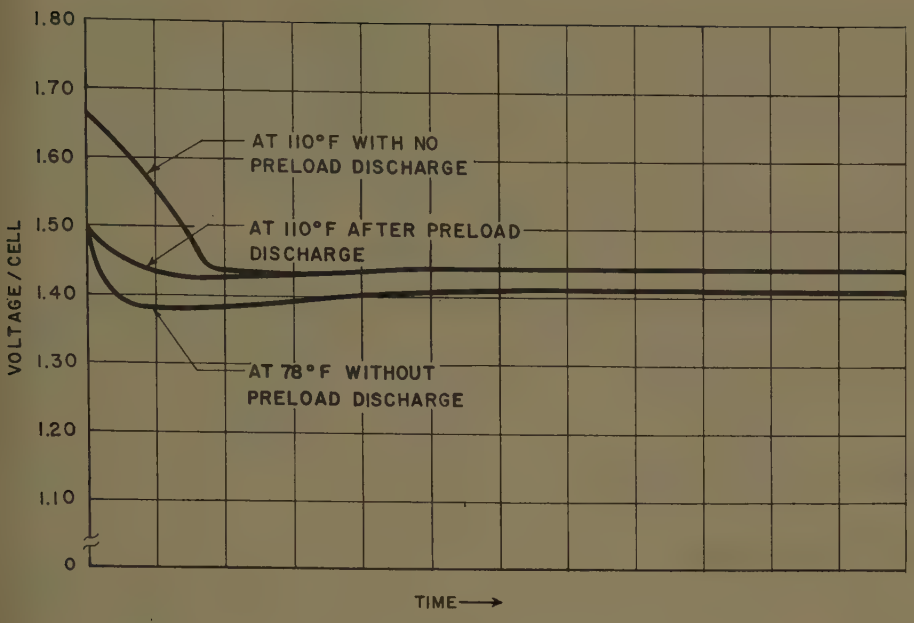


Fig. 4. Effect of temperature and preload discharge on cell voltage

jacent areas from contamination by electrolyte fumes was originally planned to be provided by completely sealing the battery case. The first test conducted under full-load conditions resulted in complete destruction of the battery. As the test progressed, the electrolyte reached its boiling point which in turn increased the internal case pressure and the electrolyte boiling point until the cell-case-material softening temperature was exceeded. An alternate scheme to exhaust the fumes overboard was then simulated in an altitude chamber. Both the voltage regulation and available ampere-hour capacity suffered severely at high altitudes. A third test finally was conducted to better define the effect of altitude. One group of cells in a bell jar was electrically connected in series to another group of cells outside the bell jar. After evacuation of the bell jar to less than 1 pound per square inch absolute, the design loads were applied and the individual cell voltages recorded. The results of the latter test are shown in Fig. 3. To provide both protection from compartment contamination by electrolyte fumes and to make full use of the battery's potentialities, a pressure-relief valve set for approximately sea-level pressure was placed in the exhaust line.

Two changes in the discharge characteristics of silver-zinc cells occur as the ambient temperature is varied. The plateau voltage increases approximately 1% for every rise of 20 F in temperature between 60 F and 120 F. In addition, and of more concern in maintenance of ±6% voltage tolerances, the shape of the discharge characteristic changes as the ambient temperature rises. Even at high-discharge rates, the higher voltage of the silver peroxide in the positive plates becomes predominate during the early part of the discharge if the ambient temperature is high. Preload discharging of the cells prior to use, if properly done, minimizes the dominance of the silver peroxide. Generally, expenditure of about 20% to 25% of the battery's capacity by the preload discharge is necessary to maintain less than ±6% voltage tolerance during the rest of the discharge. Fig. 4 shows discharges of cells at room temperature and 110 F without preload discharges, and one at 110 F after a preload discharge. To maintain voltages with the tolerances indicated in Fig. 1, the power and high-voltage batteries are given preload discharges before installation in the aircraft.

Battery Reliability

With lead-acid and other electrochemical systems, the battery's condition can be established by measurement of the short-circuit voltage or electrolyte specific gravity. These methods do not provide a satisfactory indication with silver-zinc cells since the internal resistance of silver-zinc cells is too low for a short-circuit test and its does not vary appreciably with the age of the cells. Also, the electrolyte is not available for test.

During the processing of thousands of cells, it has been found that failures do not occur until the input capacity decreases approximately 35% from what it is when the cells are new. Therefore, cell performance and reliability are maintained by using only cells which have an input capacity above a specified level. The cells are discharged and recharged at periodic intervals of less than 1 month for the small cells and up to 4 months for the large cells. The required frequency of these maintenance cycles has been found by measurement of the rate of energy loss during life tests conducted on each cell model. This method of cell maintenance has given 100% battery reliability to date.

Circuit Simplification

High-Voltage Battery Replacement

The high-voltage battery with its 320 individual cells is inherently an unreliable unit if age exceeds 60 days charged stand. To minimize the possibility of failures much painstaking maintenance is required. Although high-voltage battery reliability has been maintained at 100%, an active program for a replacement power supply was begun. First, the efforts were directed toward the use of dynamotors. Development problems of excessive broad-band noise and voltage drift could not be overcome, so the program was dropped.

With the advent of high-voltage high-power silicon diodes, development of a transformer-rectifier unit utilizing the 400-cycle-per-second inverter output became practicable. Not only is the high-voltage battery maintenance problem eliminated, but the system weight was reduced 15 pounds. The transformer-rectifier unit has the added advantage of enabling all final manufacturing testing to be done with the same high-voltage power source as that used in flight.

Alternate Power Sources

Periodic surveys and studies have been made of replacement systems for the secondary silver-zinc batteries. Included in these studies have been the ram-air turbine-driven alternator, the gas turbine-driven alternator and the primary silver-zinc battery systems. Until recently, the gas-turbine-driven alternator and conversion-unit development not only was uncertain but the development time and cost were prohibitive. With improvements in components it now appears that the development of the gas turbine and its power conversion units is feasible for future system designs.

The primary silver-zinc battery system which has been delayed by the lack of a positive automatic activation system also appears practical at this time. In many ways it is the best replacement for the present system, since it requires no main-

tenance. However, reliability is the major factor in selection of a replacement system and much statistical data will be needed to determine the reliability of the primary battery activation systems.

Conclusions

1. The silver-zinc battery system provides a power source for use in pilotless aircraft which is both light and reliable.

2. Battery life can be extended 50% through the use of nitrogen-filled cases.

3. Use of battery cases with pressure-relief valves overcomes the detrimental effect of altitude on voltage and capacity.

4. Preload discharges are required to maintain close voltage tolerances if the batteries are operated at elevated temperatures.

5. Reliable battery performance can be maintained by periodically putting the batteries through a discharge-charge cycle and discarding those batteries which fail to accept a pre-established minimum input.

6. Simplification of the power system and its maintenance is now possible through the use of a gas turbine-driven alternator or a primary silver-zinc battery system.

Reference

1. PRIMARY BATTERIES (book), G. W. Vinal. John Wiley & Sons, Inc., New York, N. Y., 1950, p. 265.

Direct-Current Metering of Large Electrochemical Pot-Lines

M. E. REAGAN
FELLOW AIEE

Synopsis: D-c metering of large aluminum reduction pot-lines represents a major problem. It is no easy task to get the required degree of accuracy in measuring powers of 100,000-kw direct current. Taking advantage of the magnetic-core improvement in materials made available by developments during World War II, the magnetic amplifier emerged as a tool for many applications. Among these was a major improvement of the metering accuracies for large electrochemical pot-lines. In this paper the many advantages are discussed and some conclusions reached as to the best fields of application.

IN THE ENDEAVOR to obtain greater operating efficiency, engineers designing equipment for the production of aluminum are constantly seeking new methods and new components which may be beneficially applied. Playing a major role are efforts to obtain more accurate and reliable d-c metering. More accurate metering means more accurate control, which in turn leads to more aluminum production per kilowatt-hour.

The d-c power supply for the modern pot-line is from rectifiers where many individual units are paralleled to supply the total pot-line current. Fig. 1 illustrates one unit of a typical rectifier pot-line diagram. The largest one at present operates at 125,000 amperes and 800 d-c volts, being fed from 24 individual rectifiers averaging about 5,200 amperes each. The metering of total current on any line necessitates some method of adding the individual currents. Totalizing shunts of a size required to operate in the main totalizing bus are difficult to build and install.

The aluminum reduction plants built during World War II were normally rated at 50,000 to 60,000 amperes and 600 to 650 volts. Rectifiers were used with individual shunts which fed individual ammeters mounted on their individual control boards. The individual control boards were often built as a composite duplex control unit which necessitated long runs of the shunt leads and requiring special calibrations. Since these leads operated at full bus potential, the duplex board and the operators and maintenance personnel were subjected to the electrical hazards involved.

For the following purposes the current transducer was produced: 1. to totalize readily the pot-line current; 2. to facilitate running long distances to the individual ammeters; and 3. to insulate properly the switchboard instruments from the power conductors.[1] However, other magnetic-amplifier types of transductors were developed which were much smaller in physical size, much lower in energy loss, had greater accuracy and the ability to read zero, and which were less expensive. The relatively small magnetic amplifier also permitted a much greater ease of installation as it did not have to surround the main bus.[2,3] Due to quality of the magnetic-core material, its accuracy and stability were greatly improved.

The same type of magnetic amplifier is used for metering of potentials. A large value of resistance is connected in series with the control winding. Here also, the insulation which isolates the instruments from the power conductors is very beneficial. The readings are usually duplicated on the duplex control board and in the pot-room. Potential magnetic-amplifiers are not only used to meter the full pot-line voltage but also for portions of the pot-line to expedite the location of trouble in individual cells.

The combination of direct voltage and current metering quantities may be used to indicate and record kilowatts. Recording instruments for d-c kilowatts, volts, and total pot-line amperes are valuable in keeping the operating efficiency at a high level. Besides providing the permanent record, recording instruments often prove valuable by registering conditions immediately prior, during and immediately after troubles of various sorts have occurred.

Totalizing current metering is an important indication of over-all pot-line performance. Seven different methods have been used and others are being proposed. These do not vary radically from each other. In the large-sized stations, the use of shunts in the main bus is not practical at present.

Scheme 1. The first scheme consists of a large transductor mounted around the main d-c bus. The direct current is measured in terms of an alternating current which is within standard instrument range. The direct current in the main bus serves as the control element for saturating the magnetic circuit. When the direct current is zero, the ammeter would read a minimum value (exciting current) but not zero. Fig. 2 shows a schematic diagram of construction and a diagram of connections. This system should perhaps be used only where in-

Paper 57-228, recommended by the AIEE Chemical Industry Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 21-25, 1957. Manuscript submitted October 15, 1956; made available for printing February 15, 1957.

M. E. REAGAN is with the Westinghouse Electric Corporation, East Pittsburgh, Pa.

The author wishes to acknowledge the co-operation and assistance of the engineering representatives of the Aluminum Company of America, the Anaconda Aluminum Company, the Kaiser Aluminum and Chemical Corporation, and the Reynolds Aluminum Company, who made this report possible.

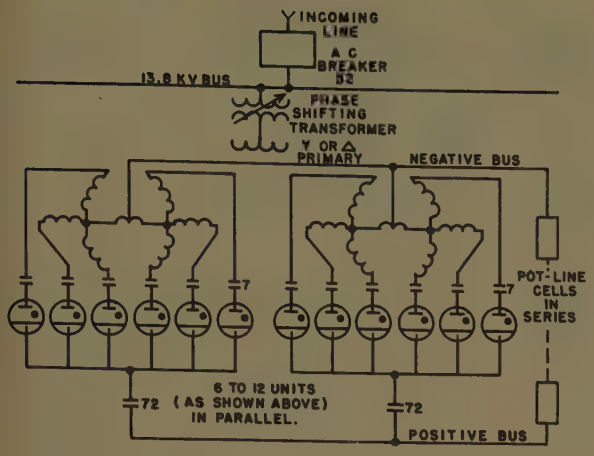


Fig. 1 (left). Schematic diagram of rectifier group feeding pot-line

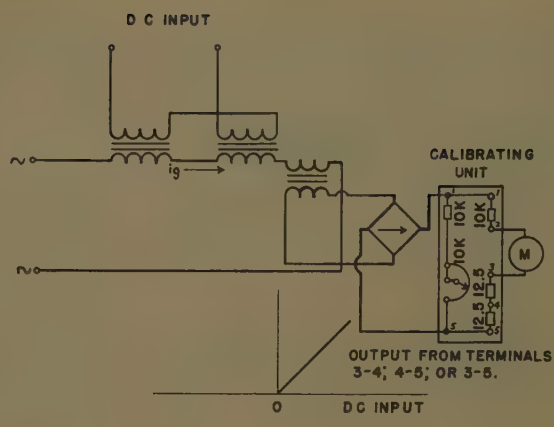


Fig. 3 (right). The input-output characteristic and schematic diagram of modern magnetic amplifier

dividual rectifier or generator outputs are not used, or as a check for a summation scheme of the individual rectifier outputs. The losses are comparatively large (the excitation for a 100,000-ampere unit being 110 amperes at 240 volts, single phase) and the first cost about ten times that of the magnetic amplifier in scheme 5. The published accuracy is 2% although field tests have shown it to perform within a 1% band of full-scale reading.

Scheme 2. The second scheme for totalizing consists of adding the outputs of a set of individual through-type transductors which measure the output of each rectifier or generator. The installed price for 12 rectifier units would be of the order of three to four times thatof scheme 1, but is necessary where individual rectifier outputs must be metered. The total excitation losses of 16 units is roughly one third of the single unit on the main d-c bus.

Scheme 3. The third totalizing scheme utilizes a d-c shunt in the cathode circuit of each rectifier (or in the positive lead of each d-c generator) which varies the control current of a magnetic amplifier with load. The output of the magnetic amplifier, in a similar fashion to scheme 2, operates the individual ammeter as well as combining with the other units to register total pot-line current; see Fig. 3. Some operators object to the use of a shunt for initiating the current indications, claiming that shunts are a source of error. This situation was thoroughly investigated and the objection was found to be true in many cases. However, it was found that the error was not due to the shunt materials or the method of manufacture but mainly to mechanical stresses caused by mounting methods. As manufactured, the shunts are very accurate, well within their guaranteed accuracy of 0.25%. When mounted in a bus, the strain on the manganin fins from the weight of the bus and the distortion resulting from bolt-tightening caused errors which have been reported as high as 20%. Another source of shunt error has been due to solder disintegration at the junction of the manganin and copper end castings. In spite of this, it is felt that the shunt is the best and most accurate means of measuring direct current when not distorted by the physical mounting. In some cases, the shunt should be supported at both ends and a flexible connector used at one end to minimize variations in output as the bus tends to move with temperature variations. Even though the shunt is changed from its factory calibration, this may be compensated for completely by changing the output of the magnetic amplifier to the correct value. There is at present no evidence that the shunt input-output ratio changes after installation.

The method of connecting the magnetic amplifier to the shunt must be studied. Due to the a-c ripple in a rectifier output, the variable field pattern induces a small voltage in the loop ordinarily resulting from the shunt lead configuration. To minimize this error, it is necessary to use flexible shunt leads and twist them as tightly as possible.

Scheme 4. The fourth method is identical to scheme 3, except that a millivolt drop is taken from the cathode or d-c bus instead of the shunt. On rectifiers, this scheme is fully as satisfactory providing the ambient temperature does not vary over too wide limits. This offers a very satisfactory scheme for those who object to the use of shunts.

Scheme 5. The fifth method uses a magnetic amplifier mounted near the main bus, using a multiple-wire tap, one on each bar of the bus. At least 6 feet of bus bar (12 feet is more desirable) must be spanned to get satisfactory results; see Fig. 4. The tapped section should not have any joints in it. The control wire (one per bar) should be placed physically close to the bus but insulated from it. One installation has the control wire embedded in a shallow saw-cut near the top center of each bus bar. The wire is insulated from the bar by varnish. This assures that the wire temperature closely follows that of the bus bar. Another installation ran the control wire between the bus bars halfway between the top and bottom of the bars. In this location, any

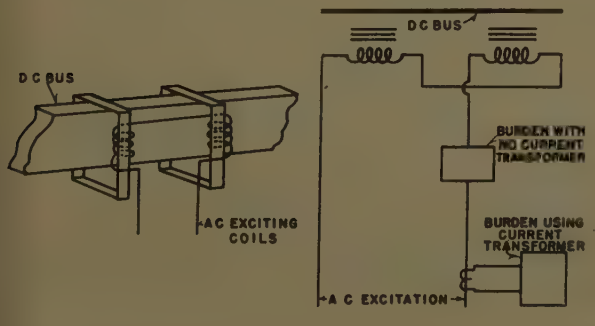


Fig. 2 (left). Through-type transductor schematic diagram, and diagram of connections

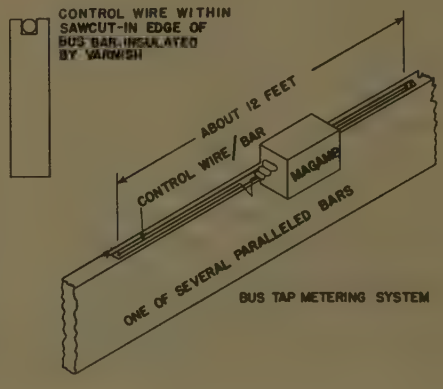


Fig. 4 (right). One of several schemes for reducing temperature error in bus tap method of totalized current metering, by burying control wire in bus bar

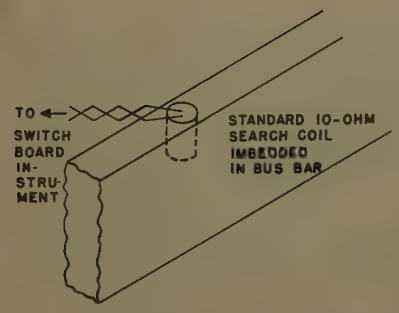


Fig. 5. Method of mounting temperature-measuring device in bus bar

nduced voltages in the control wires, by he small a-c component, is minimized.

Again referring to Fig. 3, the millivolt drop between the bus tap points will vary with the current in the bar. This millivolt drop will, with the millivolt drop of the other bus bars, determine the total magnetic-amplifier control current, and hence the magnetic-amplifier output. This is as it should be; however if the bus bar currents remain constant and the ambient temperature rises, the millivolt drop will increase, which in turn gives an error by increasing the magnetic-amplifier output. This error would be zero if the control wires increased their resistances in proportion. In fact, the error is quite small if the same grade of aluminum is used for both control wire and bus bar. The fact that the current density and ratio of current to radiating surface varies slightly between the wire and bar causes some small error. Burying the wire in the bus bar obviates the need of temperature connection under these conditions.

The output of the magnetic-amplifier tends to increase with an increase of ambient temperature with the bus current remaining constant. One suggestion for reducing this temperature error is shown in Fig. 5. To compensate fully for the error requires bus temperature measuring indicators. The manual method samples the bus temperature of several bars by means of search coils and an instrument. The operator averages the temperature readings and sets a rheostat on a calibrated scale to compensate for the change. The rheostat is electrically connected in the magnetic-amplifier totalizing circuit. The action may be made automatic by a similar connection where a 10-ohm exploring coil serves as the temperature-sensing element and, through an intermediate instrument and retransmitting rheostat, varies the millivolt drop in the current measuring circuit. If the operator notices a sudden change in the total current reading of the recording instrument, he may quickly check if the change was due to an "anode effect" in the pot-room or due to trouble in the summation of the individual rectifier outputs.

Scheme 6. A sixth method, which is very similar to scheme 5, is performing successfully; see Fig. 6. It is used when the bars of the main totalizing bus may be electrically connected together (welded) at the two tap points. This assures that there can be no potential difference between the bars due to unequal current distribution. When this is done, one large control wire (properly proportioned in size) may be used. This scheme also has the advantage of quickly recognizing an open circuit in the control wire. When multiple-control wires are used, one open circuit may temporarily be mistaken for an "anode" effect.

Scheme 7. A seventh method has been tried (in Europe) and may show some promise. It has been developed for d-c magnitudes of some 15,000 amperes. It is known as the Hall generator. The d-c bus is surrounded by an iron magnetic circuit with air gaps. Hall generators are located in the air gaps and their voltage outputs are proportional to the bus current.

Operating reports from the field installations indicate that an additional error appears under conditions when one or two rectifier units are removed for maintenance purposes. If, by tap changer adjustment, the direct voltage is brought back to its original value, the summation of current by methods 2, 3, or 4 should also be expected to return to its former level. The field tests indicate a lower current in all cases. The study of the factors which cause this are still being investigated. The field reports also state this error is not present when either of the main bus tap systems 5 or 6 is used.

In conclusion, the use of the magnetic amplifier permits some advantages:

1. They operate with greatly decreased losses when compared with the through-type bus transductor.
2. They have a high degree of accuracy, of the order of 1% or better.
3. They are static devices (with no moving parts) and therefore require minimum maintenance.
4. They are small in physical size and more easily mounted.
5. They isolate the switchboard instruments from the high-voltage d-c power busses.
6. Their use permits extremely long physical distances over which the current indications may be transmitted.

Where individual currents, as well as the totalized d-c bus current, is needed,

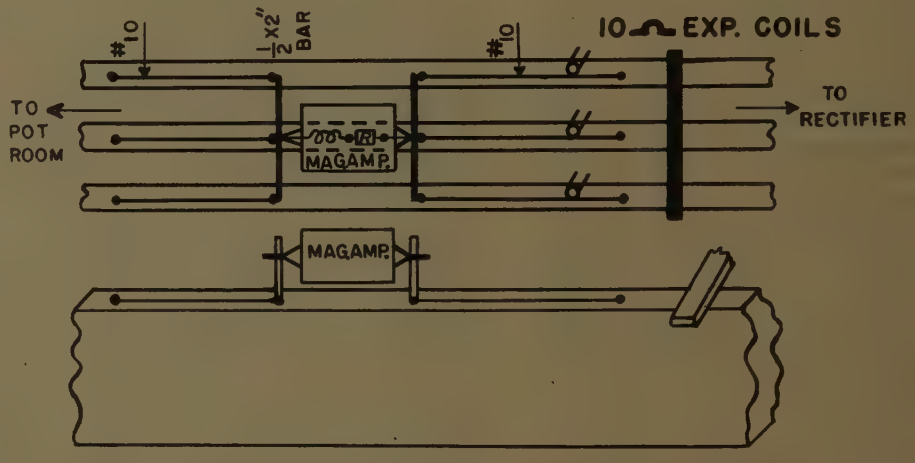


Fig. 6. Scheme of connections for bus tap system where bus bars are electrically connected on source side between tap points and source. Manganin strip R in magnetic-amplifier control winding requires use of temperature measuring and correcting equipment. Control wires are of equal length

the scheme of totalizing the shunt-controlled magnetic-amplifier outputs appears to be the second best arrangement. Up to the present, the best totalizing method appears to be that of tapping the main bus bars with magnetic-amplifier control wires.

References

1. The Theory of the Current Transductor and Its Application in the Aluminum Industry, T. R. Specht, R. N. Wagner. *AIEE Transactions*, vol. 69, pt. I, 1950, pp. 441–52.

2. Magnetic-Amplifier Applications in D-C Conversion Stations, W. A. Derr, E. J. Cham. *Ibid.*, vol. 72, pt. III, Apr. 1953, pp. 220–29.

3. Single-Ended Saturable Reactor Circuit with Quiescent Current Compensation, R. J. Radus. *Ibid.*, vol. 73, pt. I, 1954 (Feb. 1955 section), pp. 597–603.

The Response of Relay Amplifiers with Feedback

JOHN E. GIBSON
ASSOCIATE MEMBER AIEE

FRANZ B. TUTEUR
ASSOCIATE MEMBER AIEE

THE SIMPLICITY and economy of on-off controllers has led to their wide control-system application in industry. It is possible to employ the on-off controller principle in the design of a power amplifier that is simple and rugged. The use of such a device will be restricted to the operation of loads that incorporate a smoothing or low-pass effect in their response.

Basic Operation

The basic configuration of the relay amplifier is shown in Fig. 1. The relay is a polarized relay or one that closes to one contact for a positive coil voltage and to a second contact for a negative-coil voltage. A deadband or region in which neither contact is closed is typical; and usually the relay will exhibit hysteresis. That is, a larger value of coil voltage is required to close the relay than is required to hold the relay in a closed position. The filter is required for operation of the amplifier as will be seen below.

The operation of the amplifier may be shown by assuming that a step of voltage is applied to the input at t_0. Before this time the relay was open and there was no charge on the capacitor. The input voltage begins to charge the capacitor and as shown in Fig. 2, at t_1 the voltage e_r has reached a large enough value to close the relay. At t_1 then, e_m suddenly jumps from 0 to E_b. That is, the relay contact closes and the output is connected to the positive supply. The capacitor must now begin to charge to $e_{in}-E_b$ and a downward exponential begins. At t_2 the capacitor voltage has become so small that the relay cannot hold and the contact falls open. The capacitor begins to charge to e_{in} and the process continues. The output is a chain of pulses. This is certainly not a faithful reproduction of the input-voltage wave. However, if the load acts like a low-pass filter, such as a d-c motor for example, and is insensitive to frequencies as high as the pulse repetition rate, the amplifier is satisfactory and in fact, as shown below, is "linear" over most of the amplitude operating range. If the input-signal amplitude is large, the pulses are long and the average value of the output is high. The upper limit on input-signal amplitude is reached at the point at which the relay remains closed for the entire period. This limit is determined by the value of the supply voltages on the relay contacts. The lower limit on signal amplitude is determined by the value of the relay deadband.

Relay Amplifier Stability

Before discussing the input-output relations of a relay amplifier, the stability of such a device will be considered. Relays have been used directly in closed-loop systems and the design and stability criteria for relay servos have been discussed by Kochenburger in terms of a describing-function analysis.[1] The analysis of a relay amplifier may be carried on using these describing-function techniques. In Fig. 3(A) is shown a block diagram of a relay amplifier. The block H_1 represents a linear electronic amplifier and the block H_2 represents the relay, which is nonlinear. The relay static characteristic is given in Fig. 3(B). For the type of resistance adder circuit shown, a reversal of sign in H_1 or H_2 is necessary to give negative feedback.

The relay is assumed to have deadband (Δ) and hysteresis (h) as defined in Fig. 3(B). The Kochenburger describing function may be used to obtain the quasi-linear transfer function of the relay from the magnitude of a sinusoidal driving signal to the magnitude and phase angle of the fundamental of the output wave,

$$H_2=\frac{E_{m1}}{i_r}=\frac{4\sin\beta}{\pi I_r}\underline{/-\alpha} \tag{1}$$

where

$$i_r=I_r\cos\omega t \tag{2}$$

and

$$\alpha=\frac{1}{2}\left\{\cos^{-1}\frac{\Delta-h}{2I_r}-\cos^{-1}\frac{\Delta+h}{2I_r}\right\}$$
$$\beta=\frac{1}{2}\left\{\cos^{-1}\frac{\Delta-h}{2I_r}+\cos^{-1}\frac{\Delta+h}{2I_r}\right\} \tag{3}$$

The derivations of equations 1 and 3 are given by Kochenburger and require only the assumption that the higher harmonics of the output wave may be neglected. The deadband, hysteresis, and amplitude of the input signal enter into the evaluation of both α and β and, as seen below, may shape the response so as to contribute to instability.

We may also write a quasi-linear relation for A, the negative of the loop gain

$$\hat{A}=\hat{H}_1\hat{H}_2\hat{H}_3 \tag{4}$$

where H_3 is the transfer function of the resistance-capacitance network

$$H_3=\frac{R_1}{(R_1+R_2)\left[1+\dfrac{R_1R_2CS}{R_1+R_2}\right]} \tag{5}$$

The loop is stable if the roots of the characteristic equation $(1+A)$ lie in the left half of the s plane. Let us assume that $R_1=R_2$ and that H_1, the amplifier, is perfect and has a gain of K. Then,

$$H_1H_3=\frac{KR_1}{2R_1(1+R_1CS/2)}=\frac{K'}{1+TS} \tag{6}$$

Since

$$\frac{y}{x}=\frac{H_1H_2H_3}{1+H_1H_2H_3} \tag{7}$$

we may write

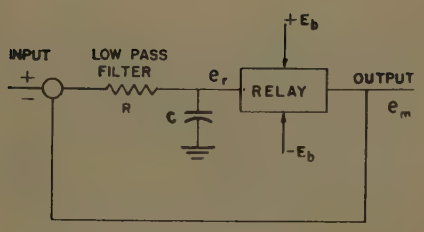
$$\frac{y}{x}=\frac{H_2}{\dfrac{1}{H_1H_3}+H_2} \tag{8}$$

According to the Routh or Leonard stability criteria, the plot of the denominator must progress around the origin keeping the origin to its left for stability as shown in Fig. 4(A). However, since H_2 depends on the amplitude of the input signal, it is simpler to plot $-H_2$ and $1/H_1H_3$ starting at the origin as in Fig. 4(B). In this figure stability is indicated by the fact that the two curves do not intersect. The frequency sensitive portion $1/H_1H_3$ is the inverse Nyquist diagram. The plot may be inverted as shown in Fig. 4(C). This is superior because the important portion of the amplitude function is expanded and the accuracy of the plot in the neighborhood of an intersection is improved. In any case the system shown in Fig. 4 is stable because there is no intersection of the loci.

In Fig. 5 is shown a specific plot for a relay amplifier with $\Delta=0.15$, $h=0.05$, and $I_{r\max}=1$. It would be assumed from

Paper 57-183, recommended by the AIEE Feedback Control Systems Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N.Y., January 21–25, 1957. Manuscript submitted August 24, 1956; made available for printing February 15, 1957.

JOHN E. GIBSON and FRANZ B. TUTEUR are with Yale University, New Haven, Conn.



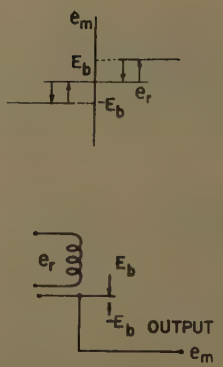


Fig. 1 (above). The basic configuration of the relay amplifier

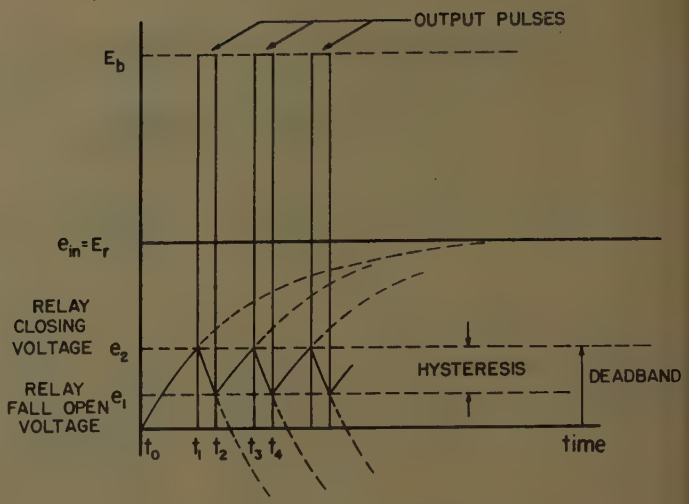


Fig. 2 (right). Voltage wave shapes in response to a step function input

the figure that under no condition of gain K or amplitude of input signal could this system be unstable. However at high gains, previously neglected lags in the amplifier or in the relay itself may cause the frequency locus to bend around as shown in Fig. 4(B), thus possibly causing instability. Fig. 5 also shows, indirectly, the effect of increasing the break frequency of the filter. If the filter-break frequency is raised to a value such that these neglected lags in the loop become important, the frequency locus will bend to the left around the origin. If it intersects the amplitude locus, instability is indicated.

Relay Amplifier Static Characteristic

The output of a relay amplifier with feedback is a series of pulses, positive or negative, of amplitude E_b and with length and repetition rate determined by the input. The output wave shape is completely independent of the load, and nowhere in the following analysis need the load be specified. Usually, however, the relay amplifier is employed in an application where only the average value of this train of pulses is important.

In order to determine the static relation between the input voltage e_r and the average output voltage E_m, we will consider H_1 and H_2, in Fig. 3, as grouped together. If the amplifier H_1 is linear, the static characteristic between e and e_m may be described as in Fig. 6. Gain in H_1 reduces directly the effective values of deadband and hysteresis, e_1 and e_2.

Let us suppose at zero time, t_0, that $e=0$. Thus the relay is open and $e_m=0$. Let a step voltage, E_r in magnitude be applied as shown in Fig. 2. Then

$$e=E_r(1-\epsilon^{-t/T_1}) \tag{9}$$

where $T_1=R_1C$. When e reaches e_2, the relay closes. Substituting e_2 for e in equation 9, we may solve for t_1, the time of closing.

$$\epsilon^{-t_1/T_1}=\frac{E_r-e_2}{E_r} \tag{10}$$

or

$$t_1=-T_1\ln(1-e_2/E_r) \tag{11}$$

After the relay closes,

$$e=E_r\left(\frac{R_2}{R_1+R_2}\right)[1-\epsilon^{-(t-t_1)/T_2}]-E_b\left(\frac{R_1}{R_1+R_2}\right)[1-\epsilon^{-(t-t_1)/T_2}]+e_2\epsilon^{-(t-t_1)/T_2} \tag{12}$$

where $T_2=(R_1R_2/R_1+R_2)C$. When e reaches e_1, the relay opens. The time of opening, t_2, may be found by setting $e=e_1$ and solving for $t=t_2$,

$$\epsilon^{-(t_2-t_1)/T_2}=\frac{e_1-E_r\left(\frac{R_2}{R_1+R_2}\right)+E_b\left(\frac{R_1}{R_1+R_2}\right)}{e_2-E_R\left(\frac{R_2}{R_1+R_2}\right)+E_b\left(\frac{R_1}{R_1+R_2}\right)} \tag{13}$$

thus

$$t_2-t_1=-T_2\times\ln\left[\frac{e_1-E_r\left(\frac{R_2}{R_1+R_2}\right)+E_b\left(\frac{R_1}{R_1+R_2}\right)}{e_2-E_r\left(\frac{R_2}{R_1+R_2}\right)+E_b\left(\frac{R_1}{R_1+R_2}\right)}\right] \tag{14}$$

Equation 15, following, is obtained when the relay is open:

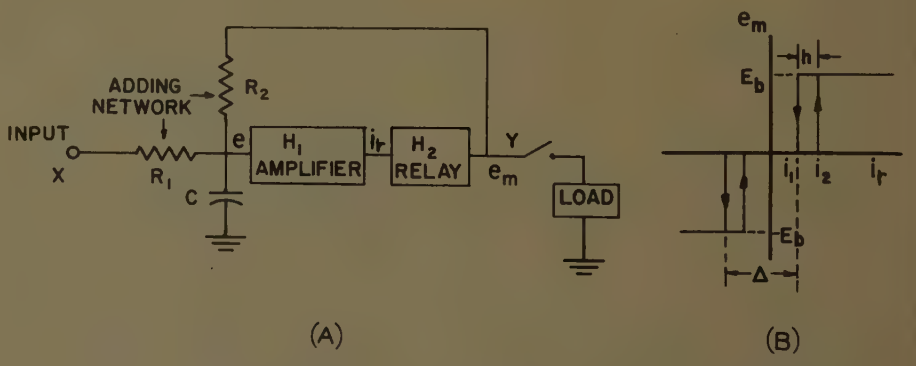


Fig. 3. A typical amplifier: (A) block diagram, (B) relay characteristics

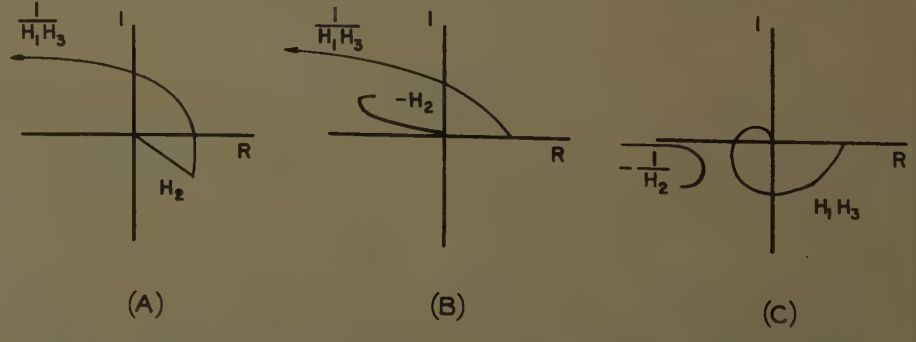


Fig. 4. Stability plots for a typical relay servo. The three plots show the same system

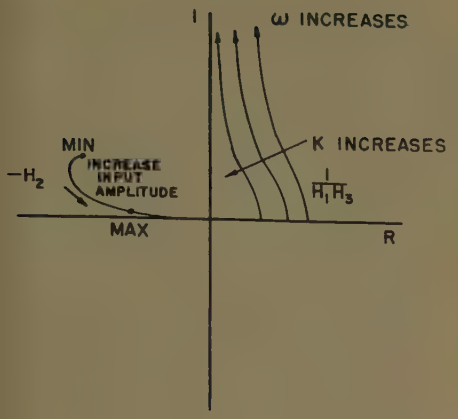


Fig. 5. Stability plot for an actual relay amplifier showing the effect of gain variation

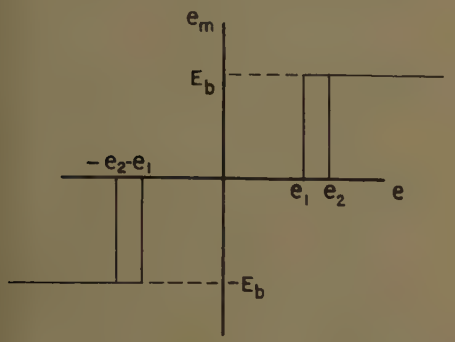


Fig. 6. Characteristic of linear amplifier and relay, deadband expressed in terms of error voltage

$$e = E_r(1-\epsilon^{-(t-t_2)/T_1}) + e_1\epsilon^{-(t-t_2)/T_1} \qquad (15)$$

It will be noted that equation 15 is the same as equation 9 plus the term due to the initial voltage on the capacitor. All of the segments of the voltage curve will be exactly the same as these described by equations 12 and 15. Equation 9 is the special case of no charge on the capacitor. The relay closes again at t_3 when $e=e_2$ given by

$$t_3-t_2 = -T_1 \ln\left(\frac{e_2-E_r}{e_1-E_r}\right) \qquad (16)$$

During the period of time (t_3-t_2) the output voltage is zero and during the period (t_2-t_1) the output is $-E_b$. These two lengths of time represent the cycle of the output-voltage wave. The average output is therefore

$$\overline{e_m} = \frac{-E_b(t_2-t_1)}{t_3-t_1} \qquad (17)$$

Substituting equations 14 and 16 into equation 17,

$$\overline{e_m} = \frac{E_b \ln \dfrac{e_1-E_r\left(\frac{R_2}{R_1+R_2}\right)+E_b\left(\frac{R_1}{R_1+R_2}\right)}{e_2-E_r\left(\frac{R_2}{R_1+R_2}\right)+E_b\left(\frac{R_1}{R_1+R_2}\right)}}{\dfrac{T_1}{T_2}\ln\dfrac{e_2-E_r}{e_1-E_r}+\ln\dfrac{e_1-E_r\left(\frac{R_2}{R_1+R_2}\right)+E_b\left(\frac{R_1}{R_1+R_2}\right)}{e_2-E_r\left(\frac{R_2}{R_1+R_2}\right)+E_b\left(\frac{R_1}{R_1+R_2}\right)}} \qquad (18)$$

By a relatively minor change in the circuit, R_2 may be grounded whenever the relay is open. This will make $T_1=T_2$ and will contribute to the linearity of the input-output relation. In this case equation 18 reduces to the following:

$$\overline{e_m} = \frac{E_b \ln\left[\dfrac{1-\left(\frac{2e_1}{E_r-E_b}\right)}{1-\left(\frac{2e_2}{E_r-E_b}\right)}\right]}{\ln\left[\dfrac{1+\frac{1}{2}\left(\frac{E_b+E_R}{e_1-E_R}\right)}{1+\frac{1}{2}\left(\frac{E_b+E_R}{e_2-E_R}\right)}\right]} \qquad (19)$$

In certain applications an electronic difference amplifier is used rather than the resistor adding network. The low-pass filter may be incorporated as a part of the interstage coupling network of the electronic amplifier. When the difference amplifier rather than the resistor adding network is employed, the factors representing voltage divider action do not appear and equation 18 may be simplified to yield

$$\overline{e_m} = \frac{E_b \ln\left[\dfrac{1-\frac{e_1}{E_r-E_b}}{1-\frac{e_2}{E_r-E_b}}\right]}{\ln\left[\dfrac{1-\frac{E_b}{E_r-e_1}}{1-\frac{E_b}{E_r-e_2}}\right]} \qquad (20)$$

By examination of this relation we can see that it gives the negative real result that is required, only for the region, $e_2 \leq E_r \leq E_b+e_1$. For $E_r<e_2$ the relay does not close: the output is zero. For $E_r>E_b+e_1$, the relay remains closed and the output is E_b. It may be shown that if e_1 and e_2 are

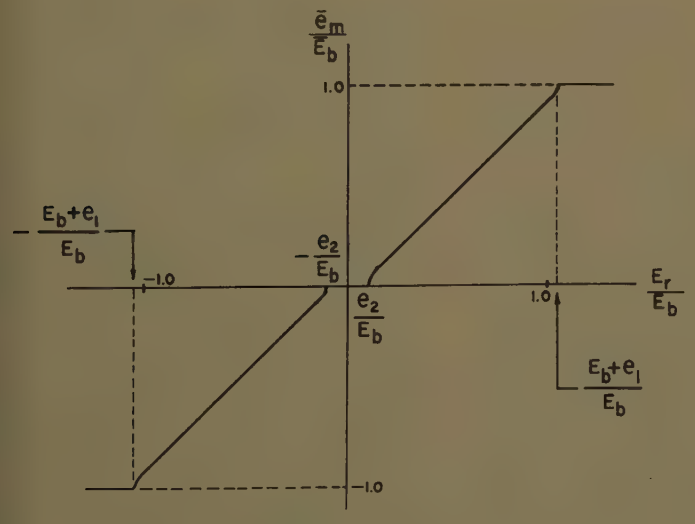


Fig. 7 (above). Static input-output characteristic of a relay amplifier

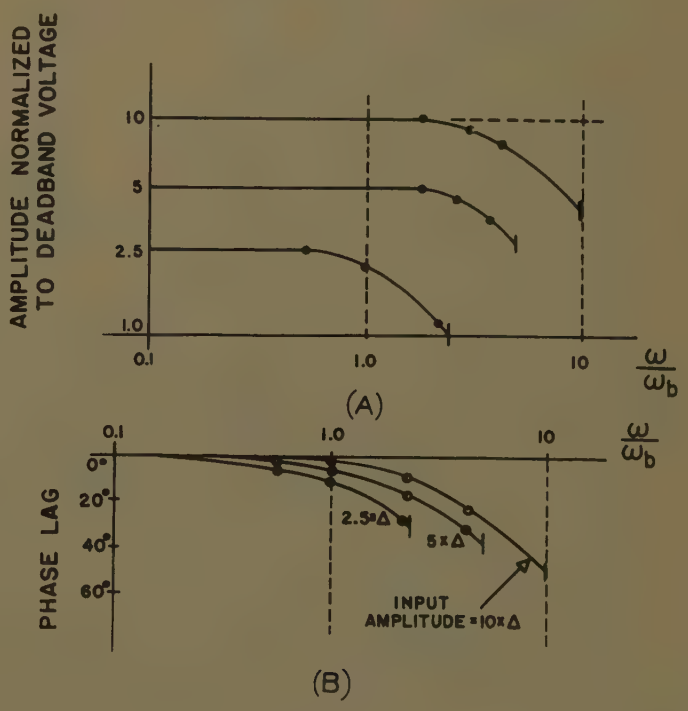
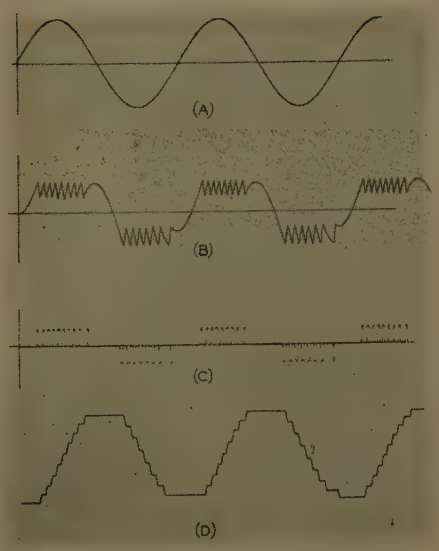


Fig. 8 (right). Frequency response of relay amplifier: (A) amplitude normalized to the value of deadband versus angular frequency to a log-log-scale; (B) phase lag versus angular frequency to a semi-log scale



small with respect to the battery voltage, the slope of e_m versus E_r is essentially constant over the operating range. The exact relation between input voltage magnitude and the average of the output voltage may be determined by the use of equation 20. In Fig. 7 is shown $\overline{e_m}$ versus E_r normalized to E_b.

Fig. 9 (left). Voltage wave shapes in response to sinusoidal driving signal: (A) sinusoidal input signal, the frequency is the same as the break frequency (ω_b) of the low-pass filter; (B) output of low-pass filter; (C) output of relay amplifier; (D) integrated output of relay amplifier

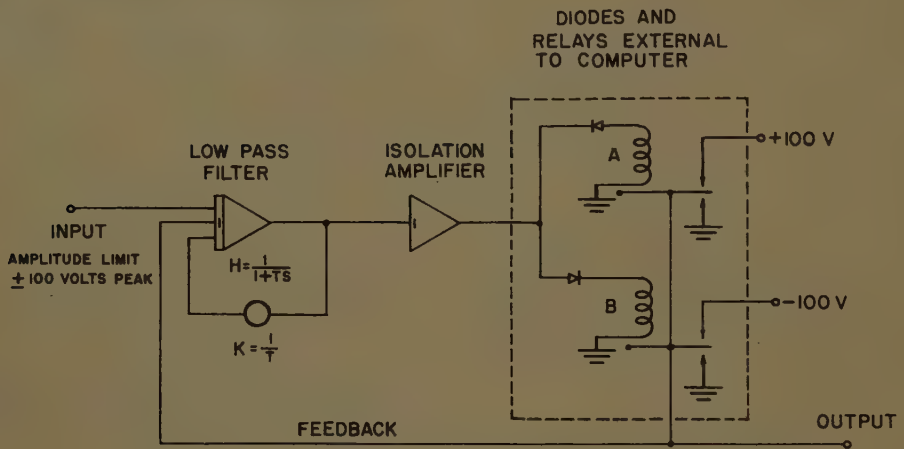


Fig. 10. Circuit for obtaining frequency response and wave shapes

Relay A: Deadband, 10.07 volts
Hysteresis, 5.17 volts

Relay B: Deadband, 9.9 volts
Hysteresis, 5.2 volts

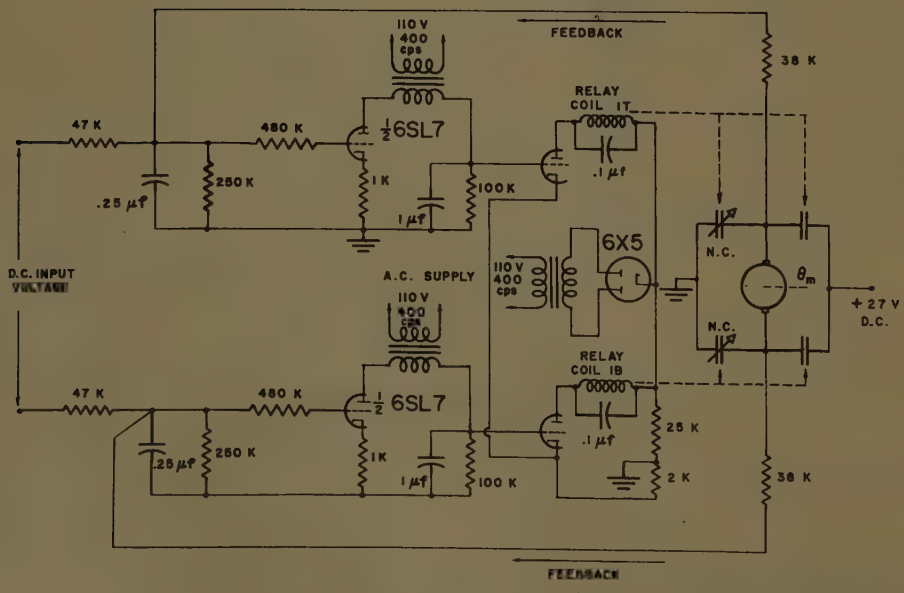


Fig. 11. A practical example of a relay amplifier used to drive the positioning motor of a servomechanism

The Frequency Response of a Relay Amplifier

The limitation on the frequency response, for a sinusoidal input, of a relay amplifier is essentially determined to a first approximation by the time constant T of the filter or the "chatter" rate of the output pulses. Actually, however, this chatter rate or output pulse repetition rate depends on the input-signal magnitude, and in the limit where E_r is either greater than E_b or zero the relay remains always closed or always open. Since the relay amplifier is a nonlinear device, the entire concept of frequency response is contrived, and a measure for one type of input signal will not hold for another. For a step function input, the relay amplifier is particularly fast and essentially establishes the required average value of $\overline{e_m}$ in one opening and closing.

The response of a relay amplifier to sinusoidal signals depends on the amplitude and the frequency of the input signal. For signals whose peak amplitude at the input of the relay is less than the value required to close the relay, there is no output. For instance, if the closing voltage of the relay in 10 volts, a signal whose peak amplitude is less than 10 volts will not close the relay. Since the low-pass filter attenuates the amplitude of the input signal, the peak amplitude must be computed at the input to the relay. At the break frequency of the filter, for instance, the filter reduces the input amplitude by a factor of 0.707, thus the minimum input signal amplitude must be 14.14 volts peak. Essentially a relay amplifier will operate properly for normal amplitude signals up to the break frequency of the low-pass filter. In Fig. 8 is shown the amplitude reduction and phase shift for a typical relay amplifier. The relay closing voltage is 0.1 of the battery voltage applied to the relay terminals and the relay drop-out voltage is 0.05 of E_b. The frequency response shown in Fig. 8(A) was obtained by smoothing the pulse train output in order to recover the input signal frequency component. When the input-signal frequency is equal to or greater than the break frequency of the filter, especially at low amplitudes, the number of output pulses per cycle becomes small and even the integrated output becomes rather step like. In Fig. 9 the various wave shapes for a particular signal are shown. In 9(A) is shown the sine-wave input to the amplifier. Its peak amplitude is 0.5 and its frequency is the same as the break frequency of the amplifier low-pass filter. In 9(B) is shown the wave form at the output

low-pass filter which illustrates the dband and hysteresis of the relay. 9(C) is given the output of the relay, l in 9(D) is shown the integrated out-. In a typical power-amplifier appli-ion the relay output drives a d-c motor ich would have an integration plus a -pass filter. Thus it could be expected t the wave shape in 9(D) would be oothed by the filter before appearing at output shaft.

The frequency response shown in Fig. 8 l the wave shapes shown in Fig. 9 were tained with the circuit shown in Fig.

The relays used were single-pole uble-throw and connected with diodes form the polarized characteristic. The -pass filter and the subtractor were nthesized on an analog computer, ich made data taking and recording of ve shapes convenient. A practical mple of a relay amplifier used to drive e actuating motors of an automatic field otter[2] is shown in Fig. 11. Here sep-ate amplifiers are used to drive the two ay coils required to complete the motor ntacts.

It is difficult to establish an equivalent op-gain constant for the relay amplifier th feedback without making several ther arbitrary assumptions. However, m Fig. 8 it can be seen that such an equivalent gain depends on the input-signal amplitude. At maximum amplitude signal the break frequency of the closed-loop response of the experimental amplifier occurs at about $4/T$ and for minimum amplitude signals the break approaches $1/T$. This would indicate a loop-gain constant of about 3 at maximum signal amplitude and zero for minimum amplitude signals.

The amount of phase shift contributed by the filter to the over-all phase shift shown in Fig. 8(B) is inversely related to the gain of the amplifier, due to the negative feedback. In addition to decreasing the closed-loop phase shift, an increase in amplifier gain will act to reduce the effective deadband, thus increasing the usefulness of the amplifier at low-signal amplitudes. As the gain is increased, the loop becomes less stable due to the increased effect of previously neglected lags in the amplifier and relay itself, until finally the amplifier goes into stable oscillation. That is, its output for no input is a continuous string of alternate positive and negative pulses of a predictable frequency. As a signal is placed on the input terminals, the length of the positive pulses grows longer and the negative pulses shorter or vice versa, thus the averaged output follows the input. This type of operation is more sensitive to low-amplitude signals than stable operation and will have minimum phase shift from input to output. However, the continuous vibration of the power element under quiescent conditions may be undesirable.

Conclusions

The static response to step function inputs to a relay amplifier with feedback has been shown to be linear over the normal operating range. The dynamic response of such amplifier has been discussed and the response to sine-wave inputs has been determined. Extension of these results to arbitrary inputs must be done with considerable care. The effects of raising the gain in the forward path of the amplifier have been discussed and the effect of relay deadband and hysteresis on the response of the amplifier has been shown.

References

1. A Frequency Response Method for Analyzing and Synthesizing Contactor Servomechanisms, **Ralph J. Kochenburger.** *AIEE Transactions*, vol. 69, pt. I, 1950, pp. 270–84.

2. An Automatic Field Plotter, **R. Gelfand, B. J. Shinn, F. B. Tuteur.** *Ibid.*, vol. 74, pt. I, Mar. 1955, pp. 73–78.

Heat Pump and Heating Cables Installed n the Same Residence for Data Purposes

C. W. JONES
MEMBER AIEE

E. E. LINDEN
MEMBER AIEE

HIS paper deals with an installation of electric heating and air conditioning a New England residence. The in-llation was made through the close -operation of the local utility and a nily building a new home, and serves e particular purpose of providing the ility with information. It is probably ique because the one house features o full-capacity systems.

A water-to-air heat pump with supple-ntary resistance heaters.

Heating cables buried in the ceiling ster.

The reason for this double system is provide information on each separately, the same time forming a close compari-n of the two in terms of the following:

1. Installation features and costs.
2. Operating and maintenance costs.
3. Effect on the utility system.
4. Promotional features.

Realizing that many utilities are active in this field, as witnessed by a recent count of some 300,000 installations, this paper is offered as a contribution toward the efforts of the newcomers and for criticism by those already in the field.

Basic Approach

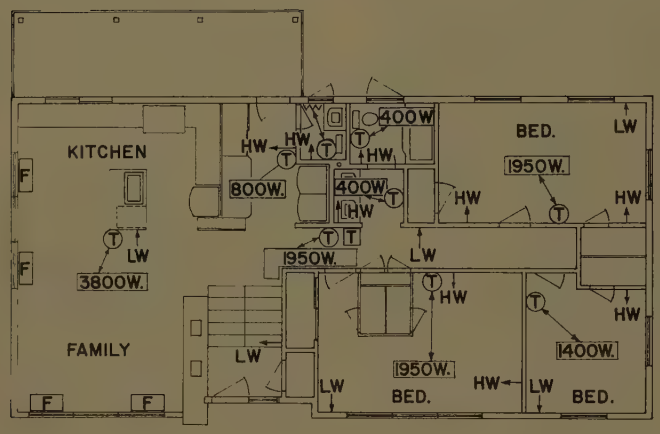
Careful thought was given to the implications in a pilot installation of this nature before planning the details. The utility involved had sponsored two experimental heat-pump installations 10 years ago and kept up to date on the subject even though the experimental heat pumps had not proved entirely suitable. With this background a list of basic notions was made, and it seemed proper that these should form the framework within which a pilot installation should be consistent. No effort is made to generalize beyond this utility's area; however, this does not mean that a similar summary would not apply elsewhere. Briefly the basic approach is shown in the following:

1. Public acceptance of electric heating and air conditioning will in large depend on the attitude of the local utility. Since electricity is the utility's most important if not sole product, perhaps any sort of negative attitude on this new load is unwise. If this is the case, the utility should then become as fully informed on the subject as

Paper 57-539, recommended by the AIEE Domestic and Commercial Applications Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE North Eastern District Meeting, Pittsfield, Mass., May 1–4, 1957. Manuscript submitted February 19, 1957; made available for printing May 23, 1957.

C. W. Jones and E. E. Linden are with the Narragansett Electric Company, Providence, R. I.

(A)



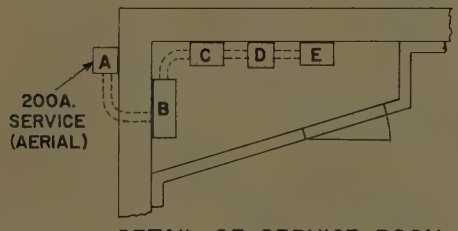
LEGEND-RADIANT CEILING

- SPACE HEATER-FAN
- Ⓣ THERMOSTAT LOCATIONS
- FEED
- 2400W. CEILING WATTAGE AT 240 V.
- BASEBOARD UNIT-300 WATTS

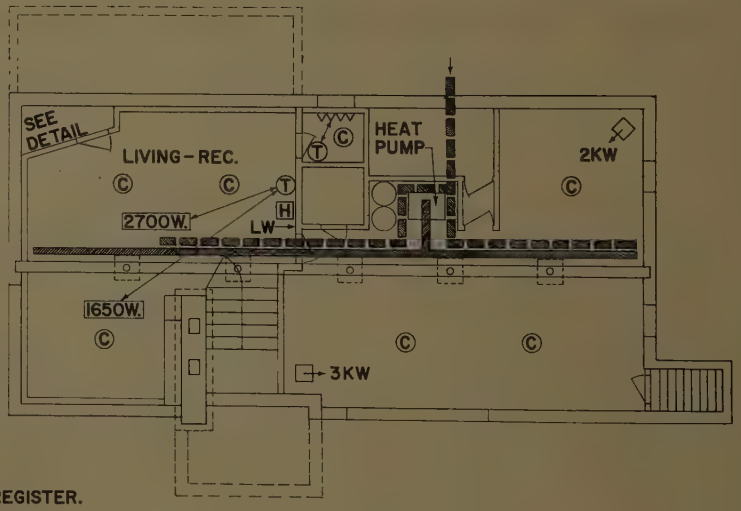
LEGEND-HEAT PUMP

- AIR SUPPLY TRUNK (NOT TO SCALE)
- AIR RETURN TRUNK (NOT TO SCALE)
- SUPPLY OR RETURN GRILLE
- Ⓒ CEILING DEFUSER
- F FLOOR DEFUSER
- HW HIGH WALL LOCATION
- LW LOW WALL LOCATION
- T THERMOSTAT CONTROLLER
- H HUMIDISTAT LOCATION

(B)



DETAIL OF SERVICE ROOM



A 2 DIAL OFF-PEAK METER.
B CHART DEMAND (15 MINUTE) TOTAL SERVICE.
C DISTRIBUTION PANEL-QUICK LAG BREAKERS.
1-50A. HEAT PUMP CIRCUIT
1-50A. DUCT HEATER CIRCUIT
OTHER DOMESTIC CIRCUITS
D WATTHOUR METER WITH DEMAND (15 MIN.) REGISTER.
E SUB DISTRIBUTION PANEL FOR RADIANT CEILING CIRCUITS.

(C)

Fig. 1. Plan of new home

A—Architect's sketch
B—First-floor plan
C—Basement plan

possible and later decide whether actively to promote or merely deal with inquiries if and when they arise.

2. Any utility policy on this subject should develop only after the results of enough local experience, so that such policy will have a valid basis. This might involve one or several pilot installations.

3. Automatic control should be fully developed in any pilot plant, whether it involves heat-pump or direct-conversion equipment. Packaged equipment should be preferred to specially fabricated equipment, the latter having little or no long-range implication.

4. Water-source heat pumps are simpler and more efficient over-all than the air-source heat pump; however, the application of water-source machines is generally limited to locations where a source of water is readily available and disposable. In suburban developments, where this is generally not the case, the air-source machine has its best application. It seems apparent that this gives the air-source machine a greater potential in numbers.

5. Direct-conversion electric heaters offer the simplest means of heating with complete zoning and low first cost; and, where there is no particular desire for cooling or dehumidification, acceptance will depend upon whether a home-owner will bear the cost of operation for the convenience, released space, etc. For supplementary heating, direct conversion has good potential.

6. The best application of either heat pump or direct conversion is in new construction, probably occurring first in homes priced $20,000 and higher. A pilot installation then should be in this price class, with both heat-pump and direct-conversion equipment, if possible, as well as other features which would be compatible with promotional use later on.

General Features of Site and Structure

Considerable time was spent in planning the details involved with the location and building features. Every effort was made to closely approximate the general features of site and construction which might be representative of the first few all-electrical homes in the area. The completion of this project allows for close study of those provisions which the utility must make in order to satisfactorily serve a load which includes full-appliance application and modern lighting as well as heating and air conditioning.

An attractive location suggests use for promotional purposes later on, however, it certainly does not detract from the all-electrical concept. The pilot installation should have an acceptable appearance to the average person and not look like a laboratory. There were favorable reactions from the various subcontractors involved. The newspaper ran a news story shortly after the structure was framed in, and later a home-section feature upon completion. As a result, the project has earned acceptance as the forerunner of things to come rather than experimentation.

Building Site, Brook, and Well

The selection of a building site was made after it was determined that a water-source heat pump would be used. A 4-acre piece of land, several miles beyond the suburbs was selected. It is well wooded with a somewhat sloped terrain and a brook. The brook is particularly attractive since it serves as a means of disposing of the water from the heat pump. Trees, for the most part, have been left to shade the house, carport, and parking area in the summer time. In winter, ample sunshine is available through the bare branches, serving to reduce the heating load on fair days.

An 8-inch well, 175 feet deep, presents a flow of 20 gpm (gallons per minute) without lowering its level beyond 65 feet. This is an unusually good well for the area, other similar wells delivering 5 to 10 gpm. Water temperature in the late winter is about 49 F (degrees Fahrenheit) and in the late summer about 51 F.

Architecture

In the architecture of the house, the resulting design would probably be termed functional-contemporary. Essentially, it is a ranch house set cornerwise into a slope. One end is for sleeping (upstairs) and storage (downstairs), and the other end (upstairs and downstairs) for living. Fig. 1(A) is a sketch of the house and Figs. 1(B) and (C) show the first-floor and basement plans. Glass is used sparingly on the bedroom end and functionally on the living-room end. The delightful feeling of openness that glass affords must be evaluated to some degree in terms of heat loss on a dark day when the wind is blowing at 30 to 40 miles per hour and the temperature is between 0 and −10 F. These are not accepted design conditions; however, such are of record. In this connection, the application of draperies which may be opened or closed in several sections reduces the heat losses through extensive glass areas as necessary, while still passing light.

The placement of living room in the basement is an attempt more fully to utilize the volume of the building for living. It is not an innovation. However, the application of a heat pump and heating cables to this type of area is certainly worthy of study.

Other Features

Insulation. The most important feature of the entire structure is the insulation. This not only reduces the heat loss and heat gain, but introduces thermal inertia which particularly minimizes the rate at which the structure cools off on cold days between heating cycles. This of course correlates to a high degree of comfort.

The economic justification of thermal insulation is well known. In this application, batt-type insulation with a vapor barrier was installed. There are 6 inches above the ceiling, 3 inches in the side walls, and 2 inches between normally heated and unheated spaces. This is essentially in accord with standards subscribed to by many manufacturers of electric heating equipment.

Windows are all double-glazed and weatherstripped. Doors to the outside are weatherstripped; storm doors have been added. Windows of the awning type in all rooms can be opened for natural ventilation on the nice days, as well as for reasons of safety.

Adequate Wiring. Wiring throughout is in accord with adequate wiring standards as a minimum.

Light Conditioning. Valence lighting when it is combined with a few functional fixtures creates an unusually pleasing effect.

Appliance Application. A full complement of electric appliances are in service.

Sizing the Heating and Air-Conditioning Equipment

Heat Loss and Gain

To size the equipment properly, heat loss and gain calculations were made by several engineers. The resulting figures differed widely on heating from 42,000 to 75,000 Btu per hour, and on cooling the difference was between 33,000 and 36,000 Btu per hour. An investigation of methods used to arrive at these figures explained the differences. A high factor of safety, a large number of air changes per hour, lack of faith in insulation values (based rightly or wrongly on the suspicion that equivalent foil insulation might be used), shading factors, and questionable design temperatures account in large for the high values. The low values are based on an almost opposite treatment. In any event this experience indicates the significance of the various judgment factors that enter into the calculation of gains and losses.

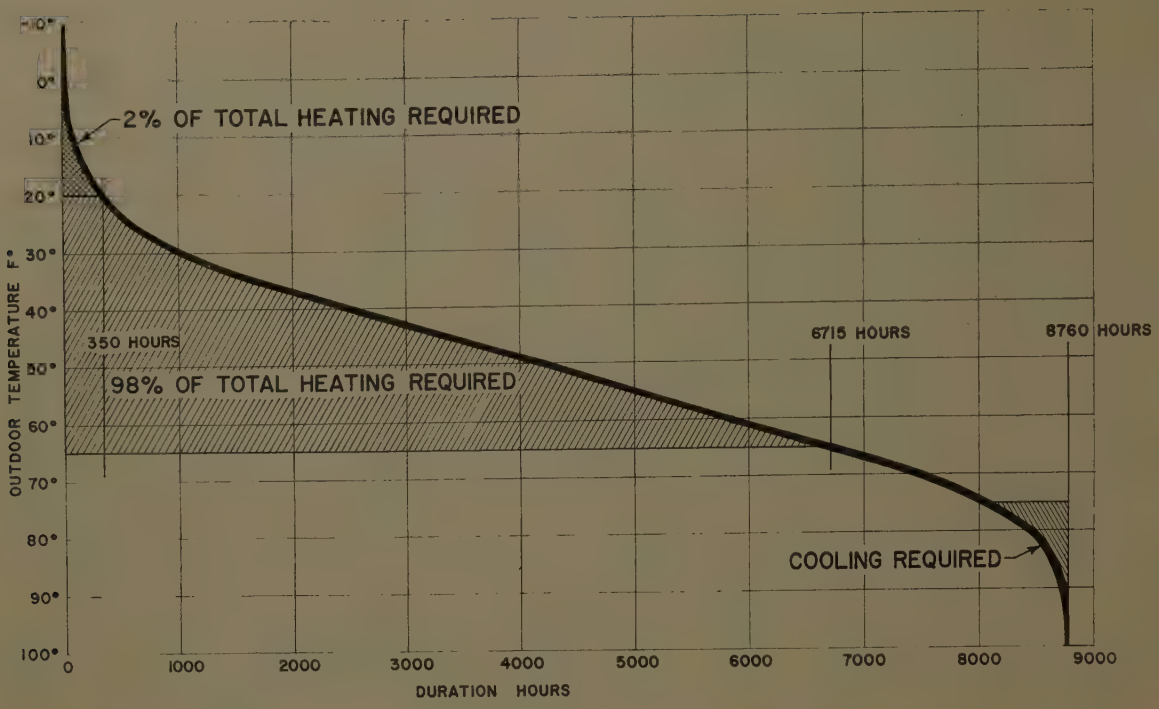


Fig. 2. Temperature duration curve

Heating-Cable Application and Selection

The heating cables buried in the ceilings have a capacity of about 60,000 Btu per hour at 240 volts and 53,000 Btu per hour at 220 volts. The manufacturer of the heating cables estimated the loss at nearer 48,000 Btu per hour and is confident that this will be proved through operation. From a practical point of view, if the installed capacity is excessive, nothing much is lost except a few installation dollars, and perhaps an occasional high, but explainable, demand. This is not serious in view of the pilot nature of the installation. On the other hand, if capacity were short, discomfort would certainly be experienced and the project could be considered a failure.

Table I. Temperature Duration, Rhode Island, July 1953 Through June 1954

Temperature, F	No. of Hours Below	Per Cent of Heating Load Below
100	8,760	
90	8,720	
80	8,450	
70	7,500	
65	6,715	100
60	5,825	81.5
50	4,175	44
40	2,475	20
35	1,660	12
30	1,025	7
25	600	4
20	350	2
15	197	1
10	71	
5	37	
0	14	
5	5	
10	0	

In this particular application, where direct-conversion equipment is involved in the same house with a heat pump, buried cables forming radiant ceilings were selected for several reasons:

1. Requires no wall space and otherwise is completely inconspicuous except for thermostats, thereby in no way restricting the arrangement of furniture or decor.
2. Represents a type of heating different in nature from the heat pump which in this case is forced air.
3. Features minimum cost which, at the same time, includes complete zoning and no expected maintenance.
4. In no way in conflict with the heat-pump installation.

Heat-Pump Application and Selection

For design purposes the forced-air system supplied by the heat pump is designed for 60,000-Btu-per-hour loss and 36,000-Btu-per-hour gain. In the selection of equipment, the water-source machine was selected because it appeared to be simpler than the air-source machines. For example, there is no "defrost cycle," and the available equipment is not so unitized in design as to almost forbid field maintenance. The water-source machine also offers better prospects in operating economies.

In the selection of size, manufacturer's specifications for a so-called 3-ton heat pump provide 36,000-Btu-per-hour heat absorption, and 42,000-Btu-per-hour on heating. A 5-ton machine provides 61,000-Btu-per-hour heat absorption and 70,000-Btu-per-hour on heating. Offhand, the 5-ton unit might seem desirable because it has more than enough capacity for both heating and cooling. However, this is not the case, because it is well known that oversizing on cooling is very undesirable. The machine will fulfill the cooling demand with short-cycled blasts of cold air, which, besides producing discomfort, may cause condensation of vapor to take place in the conditioned area rather than on the air coil of the heat pump. There is a resulting clamminess. Oversizing on heating results in similar short-cycling, which also impairs efficiency somewhat.

The conclusion reached was that, in order properly to size a heat-pump system for this climate, one of the following variations would probably be appropriate:

1. A dual heat pump, i.e., two units in tandem; one unit to operate for cooling with changeover to supply peak demand on heating, the other to operate for heating.
2. A single heat pump with direct-conversion heaters as boosters.
3. A single heat pump with storage capacity available for peak heating demand. Storage might be in water or salts.

Any of these might have been tried; however, one had to be chosen. The storage system was eliminated because of high first cost and as being too elaborate to be construed as "packaged."

Before proceeding further with the elimination, an average year was examined in order to determine the general nature of temperature duration. Referring to Table I, for a typical year there is tabulated the number of hours

in the year when the temperature was below the values listed in the "Temperature" column. Since the temperature did not exceed 100 F, for example, there are 8,760 hours when the temperature was below 100 F. In further explanation, there were 6,715 hours when the temperature was below 65 F and 350 hours when it was below 20 F.

The third column in Table I lists the percentage of annual heating load occurring below the temperatures in the first column, e.g., only 2% of the total annual heating requirement occurs below 20 F. This means that if a basic heat pump covers heating losses to 20 F, then it supplies 98% of the total Btu's used, and the tandem or supplementary unit supplies 2% of the total use; see Fig. 2.

This indicates probable few hours of use for the second unit of a tandem heat pump, similarly for any supplementary direct-conversion heaters. Furthermore, there seemed to be more reason to believe that the heat loss for the house would be under 60,000 Btu per hour rather than over.

Direct-conversion heaters are available which contain, e.g., 12 1-kw units which can be connected in any combination to give a total output of from 1 to 12 kw. It seemed reasonable, therefore, to apply the flexibility of direct-conversion heaters to take care of the admittedly nebulous difference between heat-pump capacity and the actual house loss figure.

The resulting heat-pump installation utilizes a 3-ton machine in series with a duct heater containing 12 1-kw units. The duct heaters are tentatively connected in two banks of 3 kw each, each bank being controlled by an outdoor thermostat. In addition, one 3-kw bank is used for reheat in dehumidification.

Heat-Pump and Duct-Heater Installation and Operation

The installation and operation of the heating plant designed around the heat pump is best considered part by part, beginning with the water circuit.

Water Circuit

The electrical input to the well pump must be taken into account when computing the coefficient of performance of the heat-pump system. The pump itself should be correctly sized; the plumbing should develop as little resistance to flow as possible. If, at the same time, the pump can also be used for domestic purposes, this is much better.

The water requirement for the heat pump (50 F water) is about 9 gpm on heating and 1 to 2 gpm on cooling. To take care of this, a 1/2-horsepower jet pump is used, as shown in Fig. 3. In supplying the 9 gpm, the well pump delivers directly to the heat pump. However, this direct circuit cannot be used to deliver the 1 to 2 gpm for cooling because the pump is greatly oversized for this small flow, causing a continuous series of stops and starts. Water is therefore delivered from the storage tank for the cooling requirement.

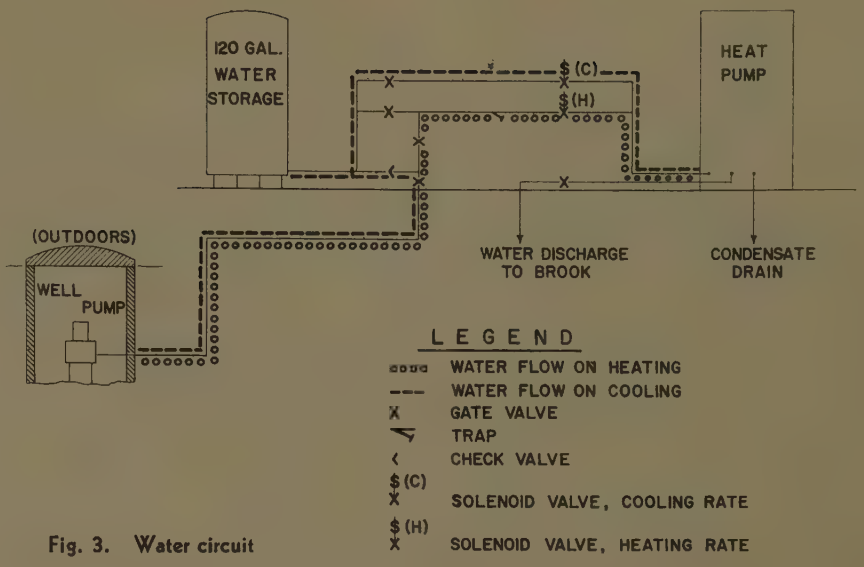


Fig. 3. Water circuit

A 120-gallon storage tank is supplied from the well pump through a check valve. This check valve prevents the storage tank from discharging into the heat pump (on heating), thereby maintaining satisfactory domestic pressure. The sizing of the storage tank is critical in the sense that its capacity must be adequate to domestic needs over any period of time that the heat pump runs. Whenever the heat pump shuts down, the well pump continues to run until storage tank pressure is recovered.

The results of delivering 9 gpm from the tank have been explored. Valves are arranged so that this can be done, although there is little practical value in this arrangement. Domestic pressure drops to below the useable range. A 1/2-horsepower pump will satisfy flow or pressure requirements in this instance, but not both.

Air Distribution

Conditioned air is distributed through a duct system designed to handle the 1,200-cubic-feet-per-minute output of the heat pump. Ample outlets and returns are provided as shown in Figs. 1(B) and (C). In the bedroom areas, supplies are located high on the inside wall with returns low on the outside wall. This provides for flexible use of wall space and keeps furniture from hampering the supply air stream. Floor diffusers supply conditioned air to the family room-kitchen area, where they are located under the large glass-wall areas. One large return is located near the floor in the center of the area. The living-recreation room is in the basement where the duct trunks run near the main beam in the ceiling. In this area, ceiling diffusers are used for supplies, with returns low in the outside walls. Outdoor air can be introduced into the return air flow. An interesting sidelight is that the duct system would have had to handle over 2,000 cubic feet per minute if a tandem heat-pump plant had been selected, increasing the cost of duct work considerably.

Heat Pump and Supplementary Duct Heaters

The electrical controls, shown schematically in Fig. 4, provide for automatic operation, including changeover from heating to cooling to dehumidification. The heat pump has a 4-way valve for automatic changeover, which is refrigerant actuated through the positioning of the throwover solenoid *RT*. A thermostat, with two heating stages *TH1* and *TH2*, one cooling stage *TC*, and a fan "auto-on" switch *CF* controls heating, cooling, and air circulation. When the first stage of heating *TH1* closes, the heating-water solenoid valve *SH* is energized and opens; at the same time, *RT* throws to heating. As soon as water is flowing through the heat pump, the compressor and circulating air fan start through the water pressure switch *WP*. The second stage of heating *TH2* provides quick pickup by switching in one 3-kw duct-heater combination

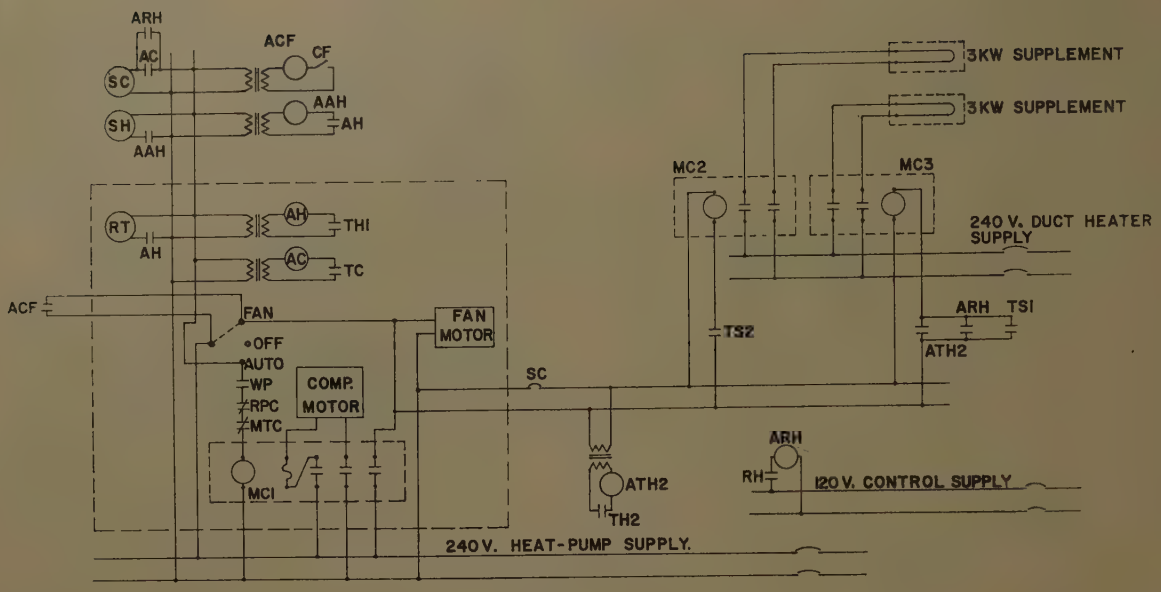


Fig. 4. Electrical schematic

LEGEND

MC1, MC2, MC3	MAGNETIC CONTACTORS	RH	RELATIVE HUMIDITY CONTROL (HUMIDISTAT)
MTC	COMPRESSOR MOTOR THERMAL CUTOUT	AH	FIRST AUXILIARY TRANSFORMER RELAY (HEATING)
RPC	REFRIGERANT PRESSURE CUTOUT	AAH	SECOND AUXILIARY TRANSFORMER RELAY (HEATING)
WP	WATER PRESSURE RELAY	AC	AUXILIARY TRANSFORMER RELAY (COOLING)
SH	SOLENOID (HEATING WATER)	ARH	AUXILIARY RELAY (RELATIVE HUMIDITY)
SC	SOLENOID (COOLING WATER)	ATH2	AUXILIARY TRANSF. RELAY (QUICK PICK-UP HEATING)
RT	REFRIGERANT THROWOVER SOLENOID	TS1	OUTDOOR THERMOSTAT, FIRST STEP SUPPLEMENTARY HEAT
TH1	INDOOR THERMOSTAT HEATING - STAGE 1	TS2	OUTDOOR THERMOSTAT, SECOND STEP SUPPLEMENTARY HEAT
TH2	INDOOR THERMOSTAT HEATING - STAGE 2	SC	SUPPLEMENTARY HEATER CUTOUT
TC	INDOOR THERMOSTAT COOLING		
CF	CIRCULATING FAN CONTROL FROM INDOOR THERMOSTAT		
ACF	AUXILIARY TRANSFORMER RELAY (CIRCULATING FAN)		

together with the heat pump. The cooling stage *TC* energizes the cooling-water solenoid *SC* valve, again starting the heat pump on water flow *WP*. Note that *RT* is normally in the cooling position. The auto-on switch *CF* is used to control the circulating air fan from upstairs so that it can be run continuously or when the compressor is running.

The heat pump as supplied is represented in Fig. 3 within the dashed lines. Included is a 3-position selector switch; "auto, "off," and "fan."

High humidity is controlled by a humidistat located in the living room; see Fig. 1(C). On a rise in relative humidity the humidistat energizes an auxiliary relay which starts the heat pump on cooling and switches in one 3-kw duct-heater combination. Return air from the room first passes over the cold coil in the heat pump, where its temperature is lowered and moisture condensed. The air then passes over the duct heater, which raises the temperature. This discharges the air to the living spaces at about the same temperature as the room return, but of course at a lower relative humidity.

It is possible for the humidistat and the thermostat to call for operation at the same time. When this happens on cooling, the operation of the system will first satisfy the setting on the humidistat before throwing over to cooling. It is not likely that heating and dehumidification will be required at the same time; however, if this should happen, then the heating requirement will first be satisfied before throwing over to dehumidification.

Two bulb-type thermostats with capillary tubes control the duct heaters on outdoor temperature, so that the electrical features can be mounted indoors and near the heat pump. The pickup setting is adjustable. To determine the temperature settings for these, the operational balance point of the heat pump is first determined.

Radiant-Ceiling Installation and Operation

The heating cables buried in plaster to form radiant ceilings are installed in accord with manufacturer's recommended procedures, except as mentioned later. In the case of two small lavettes, Figs. 1(B) and (C), the ceiling areas are too small to accommodate the smallest heating cable spaced at the 1.5 inch minimum. Baseboard heaters are used to heat these areas. Complete zoning is provided through thermostats located in each room and hallway. The storage and utility end of the basement is considered normally unheated. Space heaters are provided, however, to warm the area whenever necessary. Control of these heaters is manual.

Heating coils are supplied on reels and each wattage has a different length. A formula is used to convert the coil length and ceiling dimensions into a spacing distance. Apparently the spacing formula has a safety factor, the result being that the supply of wire was exhausted before the ceiling area in each case, causing no inconvenience.

In following the procedure for installing the cables, a staple gun was used to fasten the cables to the rock lath. Unfortunately the staple gun used was not the proper one: a wire guide had to be welded to the nose, spring tension was too great, and occasionally the gun would skip and fire with no staple. After the brown-coat plaster was applied and dried, a 600-volt Megger was used to test the insulation of each ceiling cable. In two rooms the insulation level was found to be below the 100,000 ohms specified as minimum by the manufacturer. In each case an ohmmeter was next used to determine where the insulation had been damaged. This was done by first connecting one leg of the ohmmeter to one of the leads of the heating cables at the thermostat box.

Fig. 5. Log sheets

RESIDENCE:
NAME _____
ADDRESS _____

DATE			TEMPERATURE				TIME		SET CHART	TOTAL HOUSE											REMARKS
								ELAPSED		BILLING METER KWH					SERVICE GRAPHIC						
										BLACK X 10		RED X 10			KWH X 40			KW X 40			
MONTH	DAY	YEAR	MINIMUM	MAXIMUM	MEAN	° HOURS	METER READING	HRS. MIN.		READ	USE	READ	USE	TOTAL USE	READ	DIF.	USE	READ	KW	LOAD FACT.	

DATE			HEATING PLANTS																				
			RADIANT CEILINGS						HEAT PUMP PLUS SUPPLEMENTAL I^2R														
			KWH X 20			KW X 20			HT. P. KWH		I^2R KWH		WATER PUMP					HEAT PUMP KW ①	I^2R KW ②	WATER PUMP KW ③	KW ①+②+③	L.F.
MONTH	DAY	YEAR	READ	DIF.	USE	READ	KW.	L.F.	READ	USE ①	READ	USE ②	CLOCK	HOURS USE	KWH ③	USE ①+②+③							

Next, the other leg of the ohmmeter was connected to the blade of a broad pointed screw driver. In taking subsequent ohmmeter readings, the screw driver was used to make firm contact with the brown coat by enough pressure to penetrate the brown coat approximately 1/32 inch. Readings then were recorded as the ceiling was crisscrossed in order to determine the point of least resistance, and thereby locate the fault. It was found that the point of least resistance was approximately 80% of the highest value; however, the resistance gradient adjacent to the trouble spot was great. In each case a staple driven too hard against the insulation was the cause of insulation failure, and following the foregoing procedure each was located within a radius of an inch from the point of indicated least resistance. Plastic insulating tape was used to renew the insulation. After the ceilings were finished, all ceiling cables were tested again with the Megger and found to be all right.

Electrical Service and Instrumentation

Electrical Service

An aerial 200-ampere 120/240-volt service supplies the electrical needs of

the house through facilities in the service room, shown as a detail of Fig. 1(C). Distribution and subdistribution are through quick-lag breaker panels which, beside being simple and adequate, avoid the cluttered and commercial appearance of a series of fused safety switches, or the like, all supplied through troughs. Except that special metering is involved in this installation, a service room would probably be unnecessary.

Instrumentation

The electrical service is metered through a 2-dial off-peak meter for billing purposes. In addition, the total service is chart-metered on 15-minute kilowatt demand. The ceiling subpanel, the heat pump, and the duct heaters are metered separately for kilowatt-hour and 15-minute kilowatt-hour demand (maximum indicating). The water-pump use is estimated from the hours operation of the heat pump.

Separate metering of other appliances is believed to be unnecessary, since much experience data are available on the nature of the residential load without heating. Interest is confined to the nature of each type of electric heating and its affect on the total residential demand and use. The total data indicate the impact that this type of house can be expected to have on the utility's distribution system.

In addition, various nonelectrical quantities can also be recorded:

Water Circuit. Pressure gages are located on the water inlet and discharge connections to the heat pump. Setting of the pressure switch *WP* can be checked and, in addition, water flow through the machine can be estimated from the indicated drop of pressure through the water coil.

Immersible-bulb dial-type thermometers are also located at the inlet and outlet to show the temperature change in the water coil. A stop watch and 10-gallon milk can are used to meter the flow of water at discharge.

Air Circuit. Indicating column-type thermometers are used to measure the temperature at three points: 1. return room air to the heat pump, 2. air from the heat-pump air coil, and 3. beyond the duct heaters. Temperature pickup across the heat-pump air coil and the duct heaters can thereby be recorded and compared to show an indicated coefficient of performance.

Indoor and Outdoor Air. Temperature and humidity indoors and outdoors are recorded on chart-type instruments.

Heat Pump. Refrigeration pressure gages are used to show the head and suction pressures of the refrigerant gas at the compressor. Readings from these can be combined to form an indication of heat-pump performance.

This instrumentation is adequate to show the performance of the various electrical heating features, to check the output specification of the heat pump, and to check the heat-gain and heat-loss calculations.

Records

A daily diary is kept up to date on equipment operation and performance. For convenience, some of this information is kept in log-sheet form as shown in Fig. 5.

At the writing of this paper, both heating plants have been operated satisfactorily. However, a performance report will not be released until substantial data are accumulated and evaluated.

Voltage Modulation on Aircraft Power Systems

R. E. KLOKOW
MEMBER AIEE

C. F. YOHE
ASSOCIATE MEMBER AIEE

Synopsis: This paper derives methods for calculating voltage modulation for a given speed disturbance with closed-loop voltage control. The magnitude of voltage modulation is quantitatively related to the open-loop frequency response of the voltage-regulating system.

VOLTAGE MODULATION may be defined in a broad sense as any change in the envelope of the a-c generator terminal voltage. For the case of sinusoidal voltage modulation, a precise definition is given in Appendix I.

Present and proposed military specifications are given in Appendix I. Present specifications limit voltage modulation of aircraft electric power systems to 2%. Proposed specification limits are in a range of 1.0% to 0.10%. The specifications do not explicitly state the character of voltage modulation (sinusoidal, periodic, or random), the frequency range, or the speed modulation of the generator rotor. It is hoped that the results of this paper will give information which can be used to clarify the specifications.

Since magnitudes in the order of 1.0% are involved, it is imperative that methods of calculating voltage modulation be available. It is equally important that insight into the problem of voltage modulation be gained. The causes and factors must be thoroughly investigated. Test results on actual systems are necessary, but are not sufficient to these ends. This paper presents results of part of a study of the voltage modulation problem. Before stating the precise problem considered in this paper, the broad causes of voltage modulation will be reviewed.

The generated terminal voltage is proportional to the product of flux and speed. Variations in either flux or speed will cause variations in generated voltage. Flux variations, and consequently voltage variations, may be produced by a-c generator excitation voltage changes, variation in brush contact drop (if brushes are used), nonsymmetrical generator construction, and load disturbances. It has been found from testing that the effect of variation in brush contact drop and nonsymmetrical generator construction on terminal voltage is negligible on well-designed generators. The effect of load disturbances are neglected in this paper and the generator is assumed to be operating with a balanced static load. Speed variations may be produced by torque pulsations of the prime mover, the genera-

Paper 57-470, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE East Central and Middle Eastern District Meeting and Air Transportation Conference, Dayton, Ohio, May 7–9, 1957. Manuscript submitted January 30, 1957; made available for printing May 17, 1957.

R. E. Klokow and C. F. Yohe are with the Westinghouse Electric Corporation, Lima, Ohio.

The authors wish to acknowledge the work of Prof. J. L. Klingenberger, Ohio Northern University, in measuring the open-loop transfer function of the example system, and of R. I. Daniels, mechanical design engineer, Westinghouse Electric Corporation, Lima, Ohio, for the construction of the special test equipment.

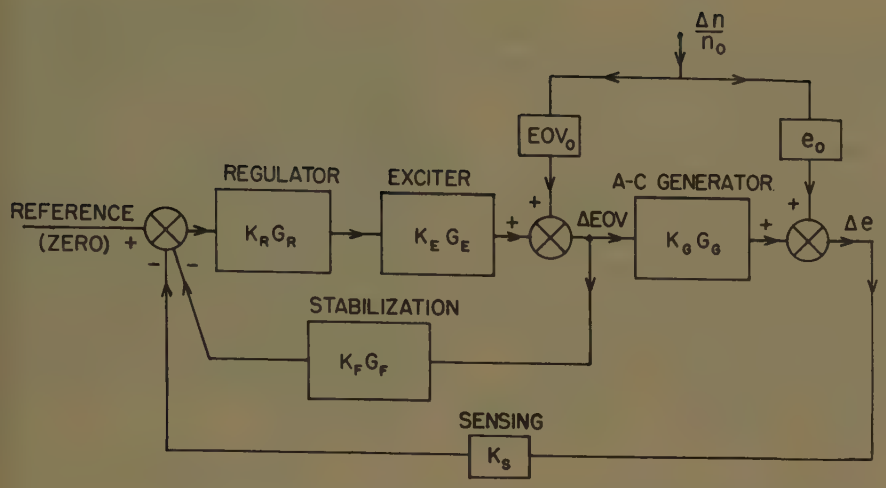


Fig. 1. Block diagram of typical voltage regulator system

tor drive, gearing, or load-switching transients. These pulsations may be amplified by the mechanical system depending on the frequency characteristics of any particular configuration. The characteristics of the mechanical system are vitally important to the problem of voltage modulation, but are not directly considered in this paper. The starting point of the study made here is with given speed modulation of the generator rotor.

Nomenclature

e = envelope of a-c generator-terminal voltage
ϕ = instantaneous flux per pole
n = instantaneous generator rotor speed (rpm)
ΔN = amplitude of sinusoidal speed modulation
f = frequency of speed changes
$\omega = 2\pi f$ = frequency of speed change in radians per second
KG = open-loop transfer function of complete system where G is a function of $(j\omega)$
$\Delta e'$ = change in e due to a-c generator speed changes
$\Delta e''$ = change in e due to exciter speed changes
Δe = instantaneous change in envelope of a-c generator terminal voltage
ΔE = maximum value of the sinusoidal change in the envelope of the a-c generator terminal voltage
K_RG_R = transfer function of regulator, where G_R is a function of $(j\omega)$
K_EG_E = transfer function of exciter, where G_E is a function of $(j\omega)$
K_GG_G = transfer function of a-c generator, where G_G is a function of $(j\omega)$
K_FG_F = transfer function of feedback circuit, where G_F is a function of $(j\omega)$
K_s = sensing circuit gain
$(EOV)_0$ = exciter output voltage at operating point
α = phase angle of flux variations with respect to speed variations

Subscript 0 indicates value of variable at operating point

Statement of Problem

The following discussion is limited to a-c generators with sinusoidal disturbances and a linearized system. However, the equations and definitions developed could be adapted for d-c generators, any given speed disturbance, and nonlinear systems. The envelope of the a-c terminal voltage is given by

$$e = K\phi n \tag{1}$$

where

ϕ = space fundamental flux per pole
n = instantaneous generator rotor speed, rpm
K = a constant of proportionality

Voltage modulation may be produced by either a flux change, a speed change, or a combined change of both variables. Equation 2, developed in Appendix II for the case of constant flux and sinusoidal variations in speed, is

$$\text{Per-cent voltage modulation} = \frac{\Delta N}{n_0} \times 100 \tag{2}$$

Therefore, with sinusoidal speed disturbances and constant a-c generator field flux, the per-cent voltage modulation is equal to the percentage speed variation.

Equation 3 developed in Appendix II for the case of constant speed and sinusoidal flux variations is

$$\text{Per-cent voltage modulation} = \frac{\Delta\Phi}{\phi_0} \times 100 \tag{3}$$

This equation states that for constant speed and sinusoidal variations in flux the percentage voltage modulation is equal to the percentage variation of flux. Equation 4 developed in Appendix II for the case of variable flux and variable speed, is as given in the following equation 4:

$$\text{Per-cent voltage modulation} = \left| \frac{\Delta N}{n_0} \underline{/0} + \frac{\Delta\Phi}{\phi_0} \underline{/\alpha} \right| \times 100 \tag{4}$$

If both speed and flux are sinusoidally varied, the per-cent voltage modulation is the magnitude of the phasor sum of the change due to each of the variables.

When an automatic voltage regulator is controlling the a-c generator terminal voltage, and speed disturbances are applied to the rotor, an error in terminal voltage will be seen by the regulator. The regulator will attempt to make corrections for the error in terminal voltage. These changes produced by the regulator will eventually be seen as changes in a-c generator field flux. Thus there is a combined variation in both speed and flux and the resulting voltage modulation will be given by equation 4. However, magnitude and phase angle of the flux change with reference to the speed change is not known since it is dependent upon the frequency response characteristics of the control system. It is the objective of this paper to derive methods of determining the resulting voltage modulation under this condition.

Assumptions

The following assumptions are made in order to expedite the discussion of this subject. However, some of these assumptions will be removed in the development of the paper:

1. The system has been linearized about no-load operation.
2. Only small excursions of the variables are considered.
3. Speed disturbances are sinusoidal functions.
4. Voltage variations due to changes in brush contact drop, load disturbances, and nonsymmetrical machine construction are neglected.
5. Only single-generator operation is considered.
6. The sinusoidal speed changes are assumed to be that of the generator rotor and exciter, and the influence of the mechanical system is not considered.
7. Effect of magnetic amplifier supply voltage variation on voltage modulation is neglected.

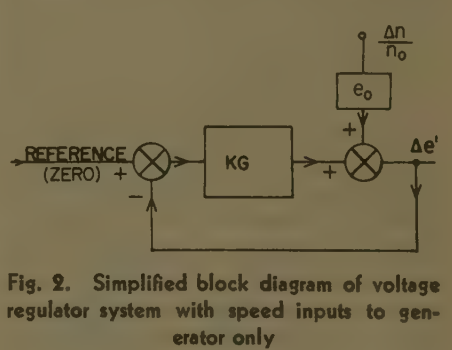


Fig. 2. Simplified block diagram of voltage regulator system with speed inputs to generator only

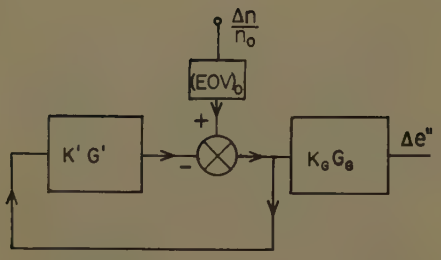


Fig. 3. Simplified block diagram of voltage regulator system with speed inputs to exciter only

Solution

Although the method presented is general, a particular system will be used in the description. Fig. 1 shows a block diagram of a typical voltage regulator system. The system is specific in the arrangement of the components but general in the sense that each component could have any transfer function. With this block diagram given, the problem reduces to that of determining

$$\frac{\Delta e/e_0}{\Delta n/n_0}$$

Since in the system there are two speed inputs, namely, to the generator and to the exciter, the combined effect due to both inputs will be determined by applying the principle of superposition. Thus, the solution will consist of two parts, that due to speed changes of the a-c generator and that due to speed changes of the exciter. The total solution will be the phasor sum of the individual solutions.

A-C Generator Speed Changes

Considering only the input to the a-c generator, the block diagram of Fig. 1 has been reduced to the block diagram shown in Fig. 2 by standard block diagram reduction techniques.[1] KG is the open-loop transfer function of the complete voltage regulator system. K is a constant independent of frequency and, for linear systems, G is a function of frequency. Thus, the component of terminal voltage change due to generator speed variations is expressed in equation 5

$$\frac{\Delta e'}{e_0} = \frac{\Delta n}{n_0}\,\frac{1}{1+KG} \qquad (5)$$

If KG is known, the voltage modulation due to the generator speed changes can be computed from equation 5 as a function of frequency. Inspection of equation 5 shows that if KG is large compared with 1, the voltage modulation is small; if KG approaches zero the voltage modulation with closed-loop control will be equal to that of a generator with constant field excitation; and when KG approaches -1 the voltage modulation is greatest but the system approaches instability.

Exciter Speed Changes

Considering inputs into the exciter only the block diagram of Fig. 1 can be reduced to that shown in Fig. 3, where $K'G'$ is defined in Appendix III. The component of change in generator terminal voltage as a function of exciter speed change is given by equation 6:

$$\frac{\Delta e''}{e_0} = \frac{\Delta n}{n_0}\,\frac{(EOV)_0}{e_0}\,K_G G_G\,\frac{1}{1+K'G'} \qquad (6)$$

Combined Speed Changes

The total change in terminal voltage due to the effect of exciter and generator speed changes is

$$\frac{\Delta e}{e_0} = \frac{\Delta e'}{e_0} + \frac{\Delta e''}{e_0} = \frac{\Delta n}{n_0}\left[\frac{1}{1+KG} + \frac{(EOV)_0}{e_0}\,K_G G_G\,\frac{1}{1+K'G'}\right] \qquad (7)$$

Expressed in terms of a ratio of per-cent voltage modulation to per-cent speed modulation, equation 7 becomes

$$\frac{\%\text{ voltage modulation}}{\%\text{ speed modulation}} = \frac{1}{1+KG} + \frac{(EOV)_0}{e_0}\,K_G G_G\,\frac{1}{1+K'G'} \qquad (8)$$

For the systems investigated by the authors the second term of equation 8, which is the voltage modulation due to the exciter speed changes has been small in comparison to the voltage modulation due to the generator speed changes. Neglecting this term, equation 8 simplifies to equation 9:

$$\frac{\%\text{ voltage modulation}}{\%\text{ speed modulation}} = \frac{1}{1+KG} \qquad (9)$$

Equation 9 shows the simplicity of the method of determining percentage voltage modulation in terms of the given percentage speed modulation. The only information required is the open-loop transfer function of the complete voltage regulator system which can be calculated or measured. This information is available in the design since it is used for determining the stability of voltage regulator systems in linear regions.

Verification of Analytical Method

Curve A of Fig. 4 represents the results of a test on a 40-kva generator system regulated by a static-type magnetic-amplifier voltage regulator. These test data were obtained by a special test fixture, as shown in Fig. 5. The a-c generator rotor was driven at a constant speed by a dynamometer. The relative speed was modulated sinusoidally by vibrating the a-c stator which was mounted in trunnion bearings. The amplitude of speed modulation was measured by providing constant excitation to the a-c generator, and measuring per-cent voltage modulation which is equal to the per-centage speed modulation. With the same speed modulation, voltage modulation was again measured with the a-c generator controlled by the regulator. The ratio of voltage modulation to speed modulation was then determined by dividing the second measurement by the first for various modulation frequencies.

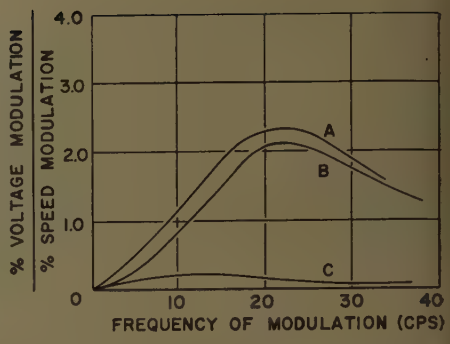


Fig. 4. Voltage modulation characteristics

Curve A—Test data
Curve B—For a-c generator speed changes only (calculated)
Curve C—For exciter speed changes only

Fig. 6 shows the open-loop frequency response characteristics of the 40-kva voltage regulator system. This data was obtained by frequency response measurements and calculations, and the information used in equation 9 to determine the voltage modulation characteristics and plotted as curve B of Fig. 4.

The computed results compare favorably with the test curve. The difference between the two curves (A and B) is due to the inaccuracy in test measurements of voltage modulation in the range of 0.1% to 1.5%, the neglect of speed input to the exciter and errors in obtaining the open-loop frequency response.

The magnitude of voltage modulation due to the speed variation of the exciter is shown as curve C of Fig. 4. This effect is seen to be small in terms of magnitudes, but has even less influence when added phasorially to curve B.

Interpretation of Results

Observation of Figs. 4 and 6 and inspection of equation 9 lead to the following interpretation:

When $|KG|$ is much greater than unity, equation 9 reduces to the following.

Fig. 5. Test fixture for modulating a-c generator speed

$$\frac{\%\text{ voltage modulation}}{\%\text{ speed modulation}} \approx \frac{1}{KG}$$

When $|KG|$ is much less than unity, equation 9 reduces to

$$\frac{\%\text{ voltage modulation}}{\%\text{ speed modulation}} \approx 1$$

When the magnitude of KG approaches -1, neither approximation holds. Fig. 6 shows that for a typical voltage regulator system the magnitude of KG is greater than unity at the low frequencies and less than unity at the high frequencies. Thus, it can be concluded that the ratio of per-cent voltage modulation to per-cent speed modulation will be less than unity at the low frequencies and equal to unity at the high frequencies. The ratio will be greater than unity at and near the natural frequency of the closed-loop voltage regulator system. The amplification in the region of the closed-loop natural frequency could reach extremely high values if the angle of KG should approach, 180 degrees when the magnitude of KG approached unity. However, for a 45-degree phase margin design, the angle of KG is held below 135 degrees when the magnitude of KG is equal to one. Consequently, for a well-designed voltage regulator system, the peak amplification can be held below an amplification of three depending upon other specification requirements, such as steady-state regulation and recovery time.

Physically, the three regions of the voltage modulation curves of Fig. 4 can be interpreted as follows: At low frequencies the regulator can change the flux in the a-c generator in the proper manner to correct for the changes in terminal voltage caused by the disturbing speed inputs. At the very high frequencies the correcting signals sensed by the voltage regulator are attenuated in magnitude to such a degree that very little change in the a-c generator flux is accomplished. Consequently at these very high frequencies the per-cent voltage modulation would be the same with or without closed-loop control. In the frequency range near the natural frequency of the closed-loop system, the regulator tries to correct for the changes in terminal voltage but the correction signal is in phase or nearly in phase with the disturbing signal because of the 180-degree shift in the sensing circuit plus the approximately 180-degree phase shift through KG. Consequently, the ratio is greater than unity in this frequency range.

Correlation of Voltage Modulation Characteristics with Other System Requirements

The voltage modulation characteristic is definitely related to the steady-state and transient performance of the voltage regulator system when subjected to load switching and speed disturbances. It is not the purpose of this paper to express these relationships quantitatively, but to give a qualitative discussion to show that the relationships exist.

Per-cent voltage regulation is equal to

$$\frac{1}{1+K} \times 100$$

K, the steady-state open-loop gain, affects voltage modulation, as can be seen from equation 9. Consequently, there is a definite relationship between voltage modulation and steady-state regulation.

Nyquist's stability criterion states that the locus of the vector $KG\ (j\omega)$ must not pass through the -1 point or encircle this point for a closed-loop system to be stable. Several textbooks show the relationship between the open-loop response $KG\ (j\omega)$ and the stability of servomechanism and regulator systems. Other authors have shown the relationship between frequency response and transient response. For purposes of this paper, it will suffice to say that there is a relationship between the stability of the system and the voltage modulation characteristic. In general, it may be said that the more oscillatory a closed-loop system is to load changes, the higher the ratio of per-cent voltage modulation to per-cent speed modulation will be at the natural frequency of the closed-loop system.

There is also a relationship between steady-state voltage regulation and stability, consequently, all three system requirements are interrelated, i.e., improvement in one (such as per-cent regulation) may result in making the other two (e.g., voltage modulation and stability) worse.

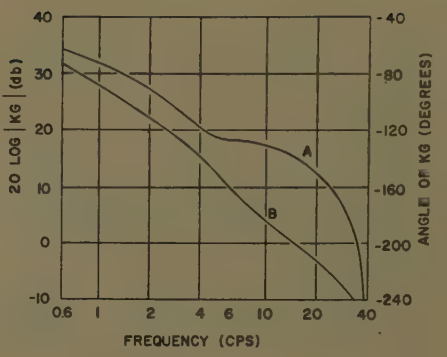


Fig. 6. Open-loop frequency response characteristics of voltage regulator system

Generalization of Method

Several assumptions were enumerated at the start of the paper. Most of these were made in order to confine the discussion and the particular system treated to a defined region. The discussion has been limited to no-load operation which is usually the most severe condition for voltage modulation. However, other load conditions could be treated by determining a new KG for the desired load.

Only small excursions of the variables were considered in order to linearize for saturation of the control system components and avoid the nonlinearity of the second-order term as given in equation 20. This second-order term will be negligible for all practical systems where the voltage modulation is small. For a given case where the speed modulation is 1.0% and the percentage change of a-c generator flux is 1.0%, to include the second-order term would not change the computed voltage modulation by more than one part in 100. Since only small excursions are encountered in the voltages of the various components for this problem, linearization about the operating point gives accurate results. Constants are computed or tested for at the operating point under consideration.

Only sinusoidal speed disturbances were treated in the development of the voltage modulation characteristics. However, the voltage modulation characteristics could be determined for nonsinusoidal but periodic waves by representing the input as a Fourier series or for a given random speed disturbance by an analog computer representation.

The effect of voltage modulation due to brush drop and nonsymmetrical machine construction was neglected. However, it has been found on all systems investigated that these effects are negligible on well-designed machines.

Only single-generator operation has been considered; however, test results

show that voltage modulation is less under parallel system operation than under single-generator operation.

It has been also assumed in this paper that the speed disturbances were those of the rotor and that this function was given. However, these speed disturbances could be determined for a given mechanical system by measuring voltage modulation with constant a-c generator excitation. This function of speed could be used in an analog computer simulation to determine voltage modulation for a particular closed-loop system.

The effect of the variation of the supply voltage to the magnetic amplifiers on voltage modulation could be evaluated by including extra loops in the block diagram. However, the calculations become lengthy and it has been found that these effects are negligible.

Conclusions

The ratio of per-cent voltage modulation to per-cent speed modulation has been quantitatively related to the open-loop frequency response characteristic of the voltage regulator system. The solution is stated mathematically in equation 9, and is discussed in a previous section. Applying sinusoidal speed disturbances to the generator of a closed-loop voltage regulator system results in the following ratio of per-cent voltage modulation to per-cent speed modulation: 1. At frequencies under the natural frequency of the closed-loop system, this ratio is less than one; 2. over natural frequency it is equal to unity; and 3. at and near the natural frequency, it is greater than one.

From the results of this paper it is apparent that per-cent voltage modulation cannot be specified independently of per-cent speed modulation, i.e., before voltage modulation can be calculated, the speed modulation must be known. On a particular system, the voltage modulation could be objectionable when the generator is driven by a given test stand and be within specification limits when driven on an aircraft. Although this paper solves the problem of determining voltage modulation with given generator rotor speed, determining voltage modulation on an actual system necessitates a study of the complete system. Cognizance must be taken of the drive, the generator mechanical system, the gearing between drive and generator, the frequency controller of the drive, and the speed or torque pulsations into the drive from the prime power source.

Since the ratio of per-cent voltage modulation to per-cent speed modulation is interrelated with the steady-state regulation of the system and the transient performance of the system, no one requirement can be looked at individually but rather they must all be considered simultaneously.

Appendix I. Definition of Voltage Modulation and Military Specifications

The following is a summary of the voltage modulation requirements from some of the military specifications. There is a trend in the later specifications to decrease the voltage modulation limits. To comply with these more stringent specification requirements, more extensive consideration must be given to the magnitude and frequency of speed variations and to the design of the closed-loop frequency characteristics of the control system. However, these specifications do not explicitly state speed variation characteristics.

Military Specifications

1. MIL-G-6099 (April 1950)

4.5.11. Output Voltage Modulation. The output voltage modulation shall be determined by means of a calibrated oscilloscope or other satisfactory means when the generator is operated at minimum, average, and maximum rated speed both at no load and full load. The per-cent modulation, computed as indicated below, with the regulator in proper operation, shall not exceed 1 per cent plus that obtained when the generator is controlled by a hand rheostat, and in no case shall exceed 2 per cent. Per-cent modulation shall be computed as follows:

$$\text{Per-cent modulation} = \frac{E_{max} - E_{min}}{E_{max} + E_{min}} \times 100$$

2. MIL-G-6099A(ASG) (November 27, 1956)

4.5.13. Output Voltage Modulation. The output voltage modulation shall be determined when the generator is operated at minimum, average, and maximum rated speeds both at no load and rated current at rated voltage. The per-cent modulation, computed as indicated below, with the regulator in proper operation shall not exceed 1 per cent. Per-cent modulation shall be computed as follows:

$$\text{Per-cent modulation} = \frac{E_{max} - E_{min}}{E_{max} + E_{min}} \times 100$$

3. MIL-G-7894A (ASG) (May 17, 1955)

3.2.1.2.6. Amplitude Modulation. The amplitude modulation of the voltage wave shall not exceed 2 per cent at any frequency.

4. XEL-907 (June 1, 1956)

3.1.3.2.5. Voltage Modulation.

3.1.3.2.5.1. Amplitude. The modulation of voltage shall not exceed an amplitude of 3.3 volts (115 volt, line-to-neutral RMS system) as measured peak-to-peak difference between the minimum voltage reached and the maximum voltage reached during any one second.

3.1.3.2.5.2. Frequency Characteristic of the Voltage Modulation. The frequency characteristics of the voltage modulation shall be confined within the envelope of limits shown in Fig. 2. [Fig. 7 of this paper]

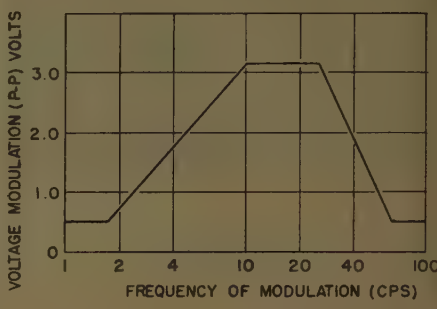


Fig. 7. Frequency characteristics for voltage modulation in a-c systems (Fig. 2, ref. 7)

Definition of Voltage Modulation

Voltage modulation has been defined in the listed military specifications in terms of percentage and peak-to-peaks volts. Per-cent voltage modulation is used throughout this paper in order to make the definition of voltage modulation independent of line-to-line, line-to-neutral, peak-to-peak, and peak volts. If consistent units of voltage are used in the following, the percentage voltage modulation computed will be correct:

$$\text{Per-cent voltage modulation} = \frac{E_{max} - E_{min}}{E_{max} + E_{min}} \times 100 \qquad (10)$$

Simplified Equation for Computing Per-Cent Voltage Modulation

The voltage E_{max} and E_{min} are graphically presented in Fig. 8 together with e_0 and ΔE for sinusoidal voltage modulation. An equivalent expression for per-cent voltage modulation in terms of e_0 and ΔE is

$$\text{Per-cent voltage modulation} = \frac{E_{max} - E_{min}}{E_{max} + E_{min}} \times 100$$

$$= \frac{(e_0 + \Delta E) - (e_0 - \Delta E)}{(e_0 + \Delta E) + (e_0 - \Delta E)} \times 100$$

$$\text{Per-cent voltage modulation} = \frac{\Delta E}{e_0} \times 100 \qquad (11)$$

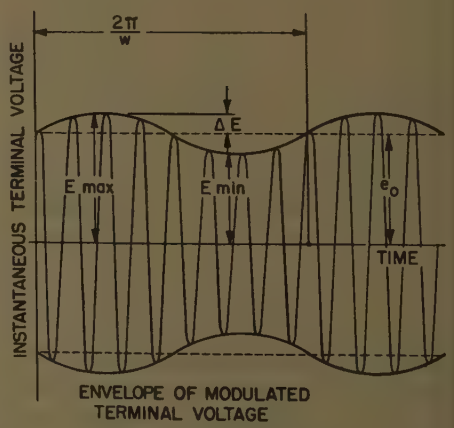


Fig. 8. Modulated a-c generator output voltage

Random-Type Voltage Modulation

The discussion in this paper has been limited to voltage modulation of a sinusoidal nature. If speed disturbances are of a random nature, the voltage modulation would be of a random nature. The present definition of voltage modulation has little meaning for the random type of speed disturbances. The voltage modulation measured from the minimum dips to the maximum peak will depend upon at what interval of time the measurements are taken and the frequency response characteristics of the metering devices. The present definition of voltage modulation will not necessarily give numerical values which are indicative of the effect of random voltage disturbances on utilization equipment.

Appendix II. Derivation of Equations

The purpose of this appendix is to derive expressions for the a-c generator terminal voltage for three combinations of generator air-gap flux and generator rotor speed. The three cases considered are

1. Constant flux and sinusoidal speed variations.
2. Constant speed and sinusoidal flux variations.
3. Sinusoidal variation of both flux and speed.

An expression for the envelope of the a-c generator terminal voltage is

$$e = K\phi n \tag{1}$$

Constant Flux and Sinusoidal Speed Variations

An expression for the change in the envelope of the a-c generator terminal voltage will be derived for the case of constant flux and sinusoidal speed variations. Let

$\phi = \phi_0$, a constant

$n = n_0 + \Delta N \sin \omega t$

$e = e_0 + \Delta e$

Substituting in equation 1 gives the desired result

$$e_0 + \Delta e = K\phi_0 \Delta N \sin \omega t + K\phi_0 n_0 \tag{12}$$

Since $e_0 = K\phi_0 n_0$

$$\Delta e = K\phi_0 \Delta N \sin \omega t \tag{13}$$

Dividing both sides of this equation by e_0 and assuming that e_0 is rated voltage gives

$$\frac{\Delta e}{e_0} = \frac{\Delta N}{n_0} \sin \omega t$$

Expressing the results in terms of the magnitude of the sinusoidal envelope gives

$$\frac{\Delta E}{e_0} = \frac{\Delta N}{n_0} \text{ or}$$

$$\text{per-cent voltage modulation} = \frac{\Delta N}{n_0} \times 100 \tag{2}$$

or, stated in words, per-cent voltage modulation equals per-cent speed modulation for the case of constant flux and sinusoidal speed perturbations.

Constant Speed and Sinusoidal Flux Changes

Let

$n = n_0$, a constant

$\phi = \phi_0 + \Delta\Phi \sin \omega t$

$e = e_0 + \Delta e$

Substitution of these relations in equation 1 yields

$$e_0 + \Delta e = Kn_0\phi_0 + Kn_0\Delta\Phi \sin \omega t \tag{14}$$

Using a similar approach to that used in the preceding section gives

$$\text{Per-cent voltage modulation} = \frac{\Delta\Phi}{\phi_0} \times 100 \tag{3}$$

Or per-cent voltage modulation equals per-cent flux modulation for the case of constant generator speed and sinusoidal flux variations.

Sinusoidal Variation of Both Flux and Speed

Let

$$n = n_0 + \Delta n \sin \omega t \tag{15}$$

$$\phi = \phi_0 + \Delta\Phi \sin (\omega t + \alpha) \tag{16}$$

Substitution into equation 1 gives

$$e_0 + \Delta e = K[\phi_0 + \Delta\Phi \sin (\omega t + \alpha)] \times (n_0 + \Delta N \sin \omega t) = K\phi_0 n_0 + K\phi_0 \Delta N \sin \omega t + Kn_0 \Delta\Phi \times \sin (\omega t + \alpha) + K\Delta\Phi\Delta N \sin (\omega t + \alpha) \sin \omega t \tag{17}$$

Neglecting the second-order term and making use of the fact that $e_0 = K\phi_0 n_0$

$$\Delta e = K\phi_0 \Delta N \sin \omega t + Kn_0 \Delta\Phi \sin (\omega t + \alpha) \tag{18}$$

Dividing both sides by e_0:

$$\frac{\Delta e}{e_0} = \frac{\Delta N}{n_0} \sin \omega t + \frac{\Delta\Phi}{\phi_0} \sin (\omega t + \alpha) \tag{19}$$

Writing equation 18 in terms of phasors gives

Per-cent voltage modulation

$$= \left| \frac{\Delta N}{n_0} \underline{|0} + \frac{\Delta\Phi}{\phi_0} \underline{|\alpha} \right| \times 100 \tag{4}$$

Equation 4 states that the amplitude of the voltage modulation envelope is equal to the magnitude of the phasor sum of per-unit speed and per-unit flux for the case of sinusoidal speed flux changes.

Appendix III. Block Diagram Representation

Linearization of Product of Flux and Speed

The block diagram given in Fig. 1 shows that terminal voltage is a function of the sum of flux and speed changes. Derivation of the equation showing this relationship follows:

From equation 1

$$e = K\phi n$$

Let: $\phi = \phi_0 + \Delta\phi$

$n = n_0 + \Delta n$

$e = e_0 + \Delta e$ (1')

Substituting these in equation 1, yields

$$e_0 + \Delta e = K(\phi_0 + \Delta\phi)(n_0 + \Delta n)$$

$$e_0 + \Delta e = K(\phi_0 n_0 + n_0\Delta\phi + \phi_0\Delta n + \Delta\phi\Delta n) \tag{20}$$

Neglecting the second-order term gives

$$\Delta e = Kn_0\Delta\phi + K\phi_0\Delta n \tag{21}$$

Dividing by e_0 gives

$$\frac{\Delta e}{e_0} = \frac{\Delta\phi}{\phi_0} + \frac{\Delta n}{n_0} \tag{22}$$

Representation of equation 22 is shown in the block diagram of Fig. 1, as is also speed input to the exciter, which was handled in a similar manner.

Incremental Variables

All variables used in the block diagrams are incremental variables. The constants are determined at the operating points of each component. For example, the gain and time constant of the generator are proportional to the slope of the generator saturation curve at the operating point. Use of incremental variables is justified in the study of voltage modulation because the changes are in the order of 1.0%. However, if the transfer characteristics are not linear over the range of variables considered, this nonlinear problem can be solved on an analog computer.

Reduction of Block Diagram

Rather than present step-by-step reduction of the block diagram shown in Fig. 1 to those given in Figs. 2 and 3, KG and $K'G'$ are defined by the following equations:

$$KG = \left(\frac{K_R G_R K_E G_E}{1 + K_F G_F K_R G_R K_E G_E} \right) K_S K_G G_G \tag{23}$$

$$K'G' = K_R G_R K_E G_E (K_F G_F + K_S K_G G_G) \tag{24}$$

References

1. Principles of Servomechanisms (book), G. S. Brown, D. P. Campbell. John Wiley & Sons, Inc., New York, N.Y., 1948.

2. Servomechanisms and Regulating System Design, Vol. I (book), H. Chestnut, R. W. M. Mayer. John Wiley & Sons, Inc., 1955.

3. Development of Definitions for Voltage Modulation Characteristics in 380- to 420-Cycle-Per-Second Aircraft Electric Systems, O. Markowitz. *AIEE Transactions*, vol. 75, pt. II, Sept. 1956, pp. 193–97.

4. General Specification for Aircraft Alternating Current Generators and Regulators. *Military Specification MIL-G-6099*, Apr. 19, 1950.

5. General Specification for Aircraft Alternating Current Generators and Regulators. *Military Specification MIL-G-6099A(ASG)*, Nov. 27, 1956.

6. Characteristics of Aircraft Electric Power. *Military Specification MIL-G-7894A (ASG)*, May 17, 1955.

7. General Specification for Alternating Current, Constant Frequency, Single Generator, Generating System. *Military Specification XEL 907*, June 1, 1956.

Design of Generating Equipment for Electra Turboprop Aircraft

J. R. M. ALGER
ASSOCIATE MEMBER AIEE

W. E. WARDEN
ASSOCIATE MEMBER AIEE

W. O. HANSEN
NONMEMBER AIEE

L. KLEIN
ASSOCIATE MEMBER AIEE

WITH the advent of the Electra turboprop aircraft, a new phase was introduced in the type of electric power generated and distributed on commercial aircraft. Previously to this time, nearly all commercial aircraft used a 28-volt d-c system. However, the requirement for higher rated systems and the requirement of the load equipment for a-c power resulted in the Electra using the 115/200-volt 3-phase Y-connected a-c system. Many models of military aircraft had already used a-c systems but the systems in production at that time mainly used rotating exciter generators with sensitive protection relays. At the time that the Electra design was initiated a static excitation system with magnetic amplifier controls and protection was under development. This system was developed in order to increase reliability and provide closer limits of system performance. This static excitation system has been applied to the Electra and this paper discusses its design features.

The system described here uses basic designs described in a paper by Plette and Butler.[1] This paper is confined to describing the designs peculiar to the Electra application which are not covered in the other papers. The particular application to the Electra provided new requirements which had not been previously encountered and which resulted in special equipment designs. The design methods described below resulted in enabling the advantages of static excitation and magnetic amplifier controls and protection to be successfully achieved.

System Description

A brief description of the system is given here in order that the relationship of the components can be understood in the discussions of their design. The system consists of a statically excited generator driven directly from the turboprop engine or a gear box mounted on the engine. This generator is air-cooled. Mounted in the generator housing is a permanent magnet generator which is used as a source of control and protective power independent of the main generator loads. The main generator consists of 3-phase windings on the stator and an 8-pole field rotor fed through slip rings. The output power from the stator is fed through the static exciter unit which is mounted in the main fuselage. The static exciter uses a voltage and current feed from the generator to supply power for excitation to the generator rotor field. Mounted near the static exciter is the voltage regulator which controls the amount of excitation power. Mounted adjacent to the voltage regulator is the control unit which supplies the control and protective functions and the necessary signals to the transfer relays.

Design Problems

The system had to meet special requirements for this particular application. The generator is electromagnetically rated at 60 kva; that is, double load is 120 kva. Lockheed Aircraft Corporation specified a generating system capable of delivering 90 kva (150% rated load) during flight from sea level to 30,000 feet when supplied cooling air per Fig. 1. The system is required to meet all load requirements at 5,700 rpm, corresponding to 380 cycles. This differs from the usual requirement for overloads at 400 cycles. The system is required to maintain 117 volts line to neutral at the end of 60 feet of feeders having a voltage drop of 3.54 volts at 150% load. The 60 feet of feeders are the connection from the outboard engine generators to the static exciters which are located in the fuselage. The voltage regulation sensing point is at the output of the exciters, which is close to the load bus. Under all environmental and operating conditions including 200% load the voltage is required to be maintained within ±4%. The voltage recovery limit requirements are shown in Fig. 2. The system is required to be capable of supplying under all operating conditions at the end of the feeders at least three per unit, 3-phase short-circuit current and at least five per unit single-phase-to-neutral short-circuit current. The control system is required to prevent power of the improper characteristics from being applied to any phase of the load equipment. This requirement is met by the use of overvoltage, undervoltage, underfrequency, and overfrequency protective devices. In addition, the system is required to protect the aircraft against damage due to short circuits on the main feeders to the aircraft structure. The control system is required to actuate the transfer relays properly so as to provide for complete automatic operation of the system. The components are required to be designed for maximum reliability and air-line serviceability and for 2,000 hours' life between overhauls. The following sections show what design features are incorporated in the components of the system in order to achieve these requirements and the results of these design features in meeting them, are presented.

Paper 57-471, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE East Central and Middle Eastern District Meeting and Air Transportation Conference, Dayton, Ohio, May 7–9, 1957. Manuscript submitted February 1, 1957; made available for printing May 16, 1957.

J. R. M. ALGER, W. E. WARDEN, W. O. HANSEN, and L. KLEIN are with the General Electric Company, Mr. Alger in Erie, Pa., Messrs. Warden and Hansen in Waynesboro, Va., and Mr. Klein in Schenectady, N. Y.

Nomenclature

$A = NrAc$

A_c = area of one rotor conductor, square inches

$a = \dfrac{1}{1 + K_s/K_r}$

$b = a/2(K_I/K_r)$

I_r = rotor current, amperes

K = constant relating N to rotor current

$K_I = KK_I'$

K_I' = constant relating iron weight to N

K_r = constant relating rotor watts to N

$K_s = \dfrac{K_s'}{K^2}$

K_s' = constant relating stator conductor weight to N

l = height of iron of rotor, inches

N = stator turns in series per phase

N_r = turns per pole in rotor

P = number of poles

R = rotor resistance, ohms

X_d = synchronous reactance, ohms

Z_0 = zero sequence impedance, ohms

Z_2 = negative sequence impedance, ohms

γ = weight density of rotor conductor material number pounds per cubic inch

ρ = resistivity of rotor conductor material $\left(\dfrac{\text{ohms} - \text{inches}^2}{\text{inch}}\right)$

ϕ_{FL} = the full-load flux kiloline

W_{as} = width of stator slot, inches

W_{ac} = width of stator conductor, inches

K_m = tooth flux factor

K_ϕ = yoke flux factor

H_{as} = height of stator slot, inches

T_c = turns per coil

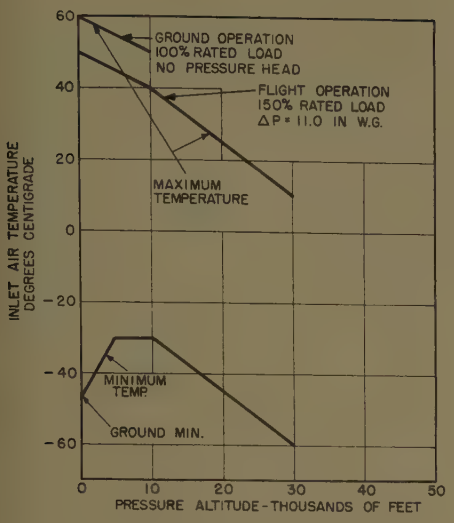


Fig. 1. Generator cooling air chart

H_{ac} = height of stator conductor, inches
A_{ac} = area of armature conductor, square mils
L_a = length of stator stack, inches
L_{cs} = length of wire in stator
Ω_{210} = resistance at operating temperature line to neutral (stator)
R_{210} = per-unit resistance at operating temperature (stator)
W = the desired stator loss, watts
W_s = computed stator loss
W_{cs} = weight of copper in stator
W_{aso} = width of armature slot opening at air gap
W_{is} = weight of iron in stator
E_{ia} = assumed per-unit induced voltage
L_{urp} = length of underside of pole face (one side)
K_{1c} = calculated value of K_1
D_{ia} = final inside diameter of stator
K_{ad} = direct axis reactance constant
K_{aq} = the quadrative axis reactance constant
X_{ad} = the direct axis reactance per unit
X_{aq} = the quadrative axis reactance, per unit
X_a = the armature leakage reactance, per unit
E_{ir} = component of induced voltage in phase with current
E_{ix} = component of induced voltage in quadrative with current
E_{ic} = the induced voltage per unit as calculated

Generator Design

Minimum Speed

The standard generators are often referred to as 380- to 420-cps (cycles per second) generators. However, the specifications normally only require overload testing at 400 cps. This application required overload operation at 380 cps (5,700 rpm) rather than the standard machine requirement for 400 cps (6,000 rpm). The decrease in rpm requires greater field current due to the decrease in induced voltage. The increase in field current is more than proportional to the speed ratio as the machine is operating under saturation at overloads. The increased field current will cause increased heating of the rotor which increases the rotor resistance. The increase in rotor resistance requires more field power and further increases the heating causing a cumulative effect. For example, a standard machine operating at 6,000 rpm and 1.5 load (90 kva) requires a field power at stabilized conditions of approximately 3 kw. To operate that machine at 5,700 rpm, 1.5 load, requires approximately 4 kw. Thus, for a 5% decrease in speed the rotor power requirements increase approximately 30%. This will overheat the rotor and also require a larger exciter. To keep the field power at 3 kw at 5,700 rpm, 1.5 load required the addition of 2 pounds of electromagnetic weight.

Overload

The voltage setting of this system is 117 volts at the end of the feeders. A ±4% tolerance is allowed on the voltage for all combinations of environmental and load conditions giving a minimum voltage of 112.2 volts. At double load the feeder voltage drop will be 4.7 volts. Therefore, the generator is designed to supply the power for the double load plus the feeder loss after the system has reached the maximum stabilized temperature for 1.5 load at 5,700 rpm. This requires power considerably in excess of the standard requirement in effect at the time the Electra design was initiated where the double-load voltage was allowed to droop without supplying any feeder power and at 6,000 rpm. This also required an increase in electromagnetic weight. Tests conducted on the Electra generator at sea level altitude, 5,700 rpm, and 57 C (degrees centigrade) inlet cooling air with an 11-inch water pressure drop across the generator showed the output voltage at the end of the feeder to be 113 volts under double load after the system had stabilized at 150% load. In the data for this test, feeder voltage drop in part was accounted for analytically, since the actual feeders were not yet installed in the test area.

Thermal

Due to the requirement for continuous operation at 5,700 rpm, 150% load, the thermal conditions for this application are actually more severe than the standard class *C* conditions. As an example, a test was made at the most severe class *C* condition of stabilized full load, 6,000 rpm at 65,000 feet altitude with −12 C inlet cooling air at 11 inches water pressure drop across the generator. Under this class *C* condition the generator operated at cooler temperatures than for the most severe Electra conditions. The thermal conditions required the application of an advanced insulation system.

The insulation system developed for this generator consists of inorganic materials all generally acceptable within AIEE class *H*. The system is the type used in military class *C* machines.

The materials used are: asbestos; glass laminates, silicone binder; glass-mica-glass laminates, silicone binder; silicone rubber; silicone varnish.

The most vulnerable component, temperature-wise, of the insulation system is the varnish. Insulation system studies using the AIEE motorette test code as a guide are presently underway.

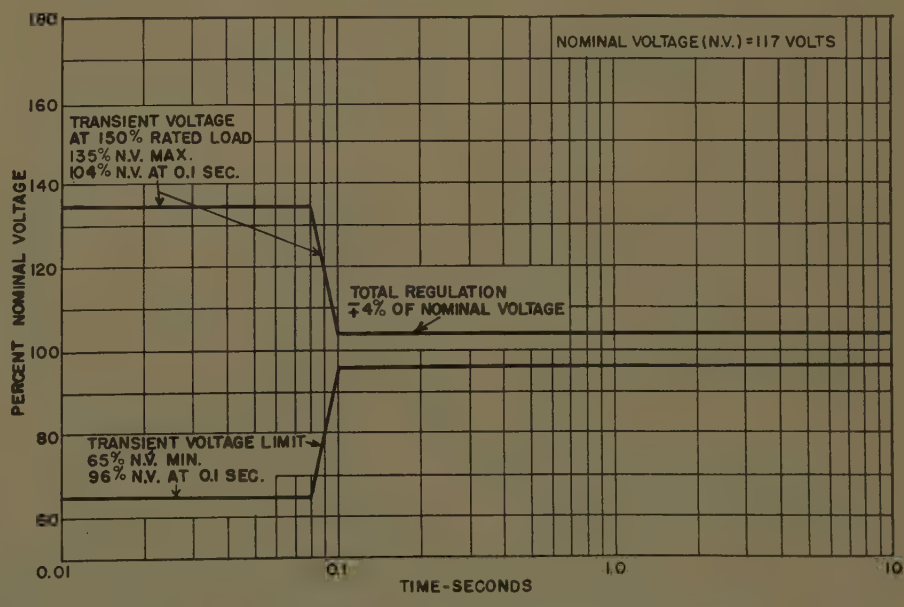


Fig. 2. Voltage recovery limits

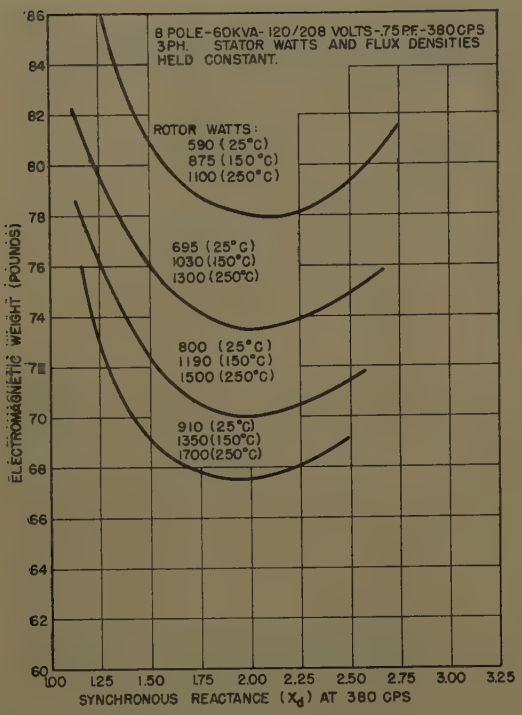


Fig. 3 (left). Electromagnetic weight versus synchronous reactance for various magnitudes of rotor watts

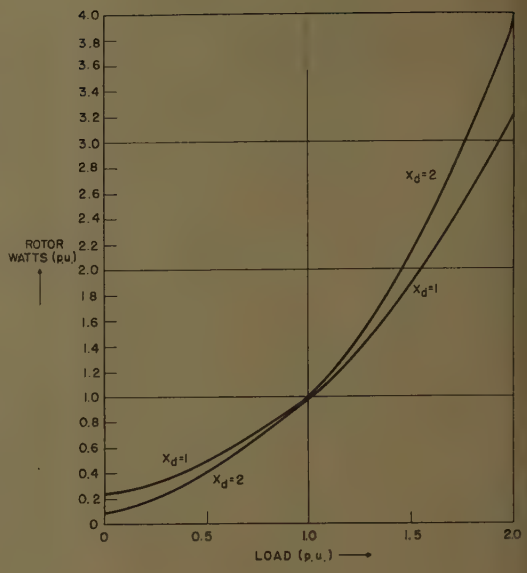


Fig. 4 (right). Rotor watts versus load

Fault Currents

All generator reactances are proportional to the square of the series turns per phase (N). Each reactance may be considered an inductance, proportional to N^2 times the permeance. Therefore, an important factor in optimizing system weight is to insure that the choice of X_d, which determines N, does not jeopardize the values of Z_0 and Z_2 needed to meet fault current or unbalanced load requirements. As a result of these factors, a small weight increase in the generator was accepted in order to keep the series turns per phase (N) at a value consistent with values of Z_0 and Z_2 that would meet the 5-per-unit single-phase fault-current requirement. It can be said that by increasing permeance, the generator reactances, particularly Z_2, may be reduced. However, an increase of stray load losses resulting in increased heating will occur if this measure is carried further than provided in the Electra design. At room ambient conditions without feeders, single-phase fault currents of over 1,200 amperes have been measured.

Generator Electrical Design

Introduction

Electrical design of an a-c aircraft generator consists mainly of choosing the proper synchronous reactance. Consistent with meeting requirements of fault current, transient response, and unbalanced load, the synchronous reactance will be chosen at the value to give minimum electromagnetic weight. For comparable designs having the same losses and saturation level, the generator electromagnetic weight varies with synchronous reactance X_d as in Fig. 3. That is, there is an optimum value below and above which the electromagnetic weight increases. For low synchronous reactance, the weight is primarily in iron while at high X_d a larger proportion of the weight is that of the conductors. A further consideration is the electromagnetic weight of the exciter. The exciter saturation level, and to some extent, exciter conductor weight, is determined by the overload excitation watts required. Therefore, exciter electromagnetic weight is considerably a function of the overloads required. A curve which presents in idealized fashion how generator rotor watts vary with load and synchronous reactance is shown in Fig. 4.

The curve shows equal watts at the continuous load point because that point determines generator life (i.e., insulation temperatures). This will require nearly equal watts at the thermally limiting point although, more exactly, the watts per square inch of conducting surface must be equal. Note that as synchronous reactance is increased, the overload watts are increased, so that even though generator electromagnetic weight may be decreased, exciter weight is increased. Therefore, the lightest weight generator-exciter combination has a synchronous reactance somewhat less than that at the minimum point on the curve of generator electromagnetic weight versus X_d.

The discussion following illustrates in simplified but ready terms why generator electromagnetic weight varies with synchronous reactance and how the digital computer is used to provide basic data to design lightweight a-c aircraft generators.

Weight Versus Synchronous Reactance

The following analysis applies to designs holding a fixed speed, a given number of poles, and the same design voltage.

The electromagnetics of an aircraft generator consist of conductor and iron weight. Minimum electromagnetic weight can be obtained for a-c generators by finding the proper weight ratio between conductor and iron. This minimum weight ratio can be described by the value of synchronous reactance, as the simplified analysis to follow will show.

The stator conductors are wound in slots with a certain number of turns (called N) in series per phase. The weight of those turns can be described by the equation:

$$\text{Stator conductor weight} = K_s' N^2$$

where K_s' is a constant. This is so because, if the number of turns in series per phase is doubled, which approximately doubles the weight, the wire area also must double, which also approximately doubles the weight, in order to maintain the same copper loss as before doubling N. In order to compare designs for weight fairly, copper losses are maintained the same. Of course, this is simplified because actually what must be done is to compare designs for weight that cool to the same temperature (i.e., have constant watts per square inch and comparable coolant velocities).

Now, how does rotor conductor weight

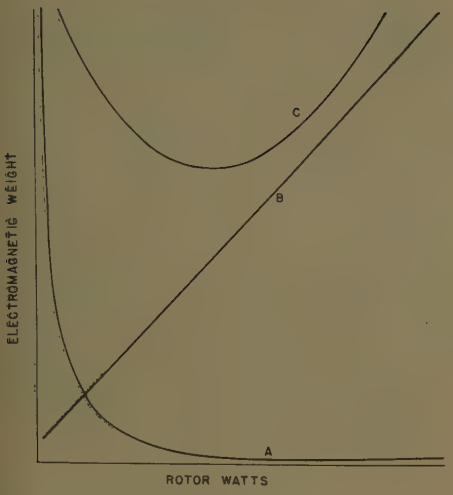


Fig. 5. System electromagnetic weight versus rotor watts

A—Generator
B—Exciter
C—Sum of generator and exciter

vary with N to maintain constant copper loss? As N is varied, the rotor current as shown in the following varies approximately as $1+KN$ for a 0.75 power factor design where K is a constant. For constant terminal voltage (strictly true only for rotating exciter machines) the full-load rotor ampere-turns (FLAT) are nearly the sum of ampere-turns required at no-load rated voltage and the ampere-turns required to supply rated current to a 3-phase short circuit.

By holding constant air-gap flux density, which is usual practice, the no-load rotor current is relatively independent of N. However, as N is varied, the ampere-turns on 3-phase short circuit vary likewise for constant output current. Therefore, in per-unit turns, the rated-load rotor ampere-turns vary as $1+KN$:

$$\text{FLAT} \propto (1+KN)$$

Rotor watts vary as rotor current squared as explained in the following and so rotor wire weight varies as $(1+KN)^2$ in order to maintain constant rotor copper loss. Rotor watts may be written as:

$$P_r = I_r^2 R = \left(\frac{\text{FLAT}}{N_r}\right)^2 \left(\frac{\rho l P N_r}{A_c}\right)$$

$$= (\text{FLAT})^2 \left(\frac{\rho l P}{A}\right)$$

where

$$A = N_r A_c$$

It is seen that rotor power is inversely proportional to "window" wire area and independent of the number of rotor turns (nelgecting insulation space). Then, since wire weight is

$$\text{Rotor conductor weight} = \gamma l A_c N_r P = \gamma l A P$$

Therefore,

$$A = \frac{\text{weight}}{\gamma l P}$$

by substitution:

$$P_r = (\text{FLAT})^2 \frac{\gamma \rho (lP)^2}{\text{weight}}$$

or for constant rotor watts:

$$\text{Rotor conductor weight} = K_r(1+KN)^2$$

where K_r is a constant, assuming relatively small length changes.

Finally, the sum of stator and rotor iron weight will vary inversely as N because comparable designs must have the same design flux densities, or otherwise saturation differences will obscure the weight comparison:

$$\text{Iron weight} = K_I'/N$$

where K_I' is another constant.

Putting the weight variations together:

$$\text{Electromagnetic weight} = K_s' N^2 + K_r(1+KN)^2 + \frac{K_I'}{N}$$

Next, it is desired to express weight variation in terms of synchronous reactance. By definition:

$$X_d = \frac{\text{Rotor amperes for rated current 3-phase short circuit}}{\text{Rotor amperes on air-gap line for rated voltage}}$$

and so, neglecting saturation at rated load, $X_d = KN$, because, as noted previously, rotor amperes for no-load rated voltage are relatively independent of N while the rated current short-circuit rotor amperes increase directly as N. Then:

$$\text{Electromotive weight} = K_s X_d^2 + K_r(1+X_d)^2 + K_I/X_d$$

where

$$K_s = \frac{K_s'}{K^2}$$

Table I. Generator Performance Versus Air Gap at 1½ Load, 5,700 Rpm, 20,000 Feet, 10 C Air

Air-Gap Length	Rotor Amperes	Rotor Watts	Rotor Rise, C
0.025	49.1	2,670	256
0.030	50.8	2,660	222

and

$$K_I = KK_I'$$

This equation shows that as X_d is raised, the iron weight decreases and the conductor weight increases. Clearly an optimum X_d for minimum weight can be obtained because a point is reached beyond which conductor weight will increase faster than iron weight is decreased. The minimum weight X_d is found where the first derivation of electromagnetic weights with respect to X_d is zero.

This is, then, the real positive root solution to the following equation (see the appendix):

$$X_d^2(X_d+a) = b$$

where

$$a = \frac{1}{1+K_s/K_r}$$

$$b = \frac{a}{2}\left(\frac{K_I}{K_r}\right)$$

Experience shows that the values of a are small compared to b. Then the solution is

$$X_d = b^{1/3}\left(1 - \frac{a}{3b^{2/3}}\right)$$

The term in parentheses is usually about 0.01 to 0.03. Neglecting it:

$$\text{Minimum weight } X_d = b^{1/3} = \left[\frac{K_I}{2(K_s+K_r)}\right]^{1/3}$$

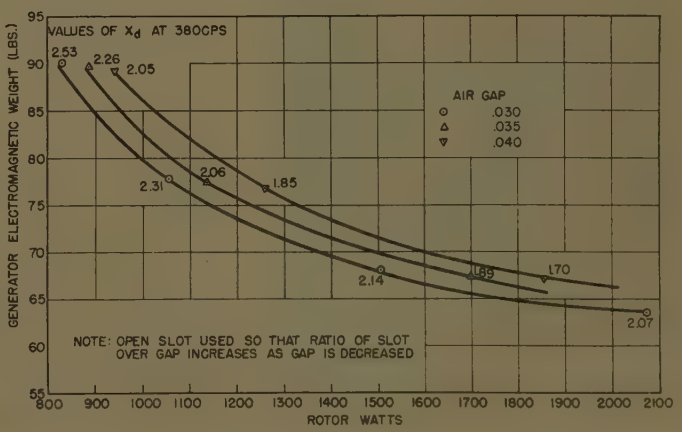


Fig. 6. Generator electromagnetic weight versus rotor watts at 250 C for various air gap lengths

The values for the Electra generator are

$X_d = 2.0$ per unit
Stator copper weight = 5.8 pounds
Rotor copper weight = 8.1 pounds
Iron weight = 49.0 pounds
Amortisseur = 2.0 pounds

For these values, the calculated constants are

$K_s = 1.45$
$K_r = 0.90$
$K_I = 98$

Putting these back into the electromagnetic weight equation, and letting X_d be an independent variable:

Electromotive weight

$$= 1.45X_d^2 + 0.9(1+X_d)^2 + \frac{98}{X_d}$$

$$= 2.35X_d^2 + 1.8X_d + \frac{98}{X_d}$$

Where this is plotted as Fig. 3, then the minimum weight X_d is

$$X_d = \left(\frac{98}{4.70}\right)^{1/3} = 2.75$$

If saturation were considered, a lower value of X_d would have resulted due to the extra iron weight necessary to maintain acceptable flux densities. Also, the static exciter and stator leakage impedance require additional voltage that again adds to the iron requirements in order that rotor power be maintained at a reasonable value. Only a complete design method can include all the factors that affect the electromagnetic weight. But, from this rough derivation, the conclusion may be drawn that the ratio of iron to copper weight determines the total electromagnetic weight; and, giving due consideration to system requirements, a minimum weight may be achieved.

Digital Computer Design Synthesis

Program Results

As an interim step in the design synthesis of lightest weight aircraft electric systems, an International Business Machines *650* computer program for generator design synthesis was developed. The program provides the data necessary to plot synchronous reactance and rotor watts versus electromagnetic weight. This points out the design for lightest weight generator; and, because the designs are relatively comparable, exciter weight can be computed and added to find the lightest weight system. See Fig. 5.

Basic results from the computer for the Electra generator are shown in Fig. 3. The curve presented shows that minimum electromagnetic weight occurs at about a synchronous reactance of 2.00. It is interesting to note that if the air-gap length is varied assuming constant fringing, the synchronous reactance will vary inversely. For a small air gap very high X_d will result. However, a given design will have less rotor copper loss at all loads if the air gap is shorter, although X_d is higher. Fig. 6 shows a calculated weight and X_d variation with air gap. It is easy to see that the lightest weight design would have zero length air gap. In that case, the rotor would not have to supply air-gap energy and could devote it all to overcoming iron saturation. Correspondingly, for a zero-length air gap, X_d would be infinite, by definition, although the rotor watts would be the least at any load. Naturally, therefore, the aircraft generator designer holds his air-gap length to the smallest possible, considering four other items:

1. Unbalanced magnetic pull.
2. Stray load losses.
3. Harmonics in the voltage.
4. Production tolerances.

Shortened air gaps require increasingly close production tolerances in order to avoid gap eccentricity. The minimum allowable gap from that standpoint is the one that lightweight mechanical construction will just still hold in place and with which the factory organization can maintain reasonable concentricity. Furthermore, shortened air gaps make relatively more difficult the task of holding down flux pulsations across the pole faces due to poorer fringing. These flux pulsations give rise to extra losses in the iron surfaces and cause extra rotor heating. It is interesting to note that because of this condition a given design will have an "optimum" air gap for minimum rotor watts, as the data in Table I illustrate.

For smaller air gaps, iron heating will cause more rotor losses than the reduced rotor current can overcome. At larger air gaps the increasing rotor current increases the losses. Of course, this is based on an existing design where fringing cannot be improved while decreasing the gap.

There are many factors to investigate

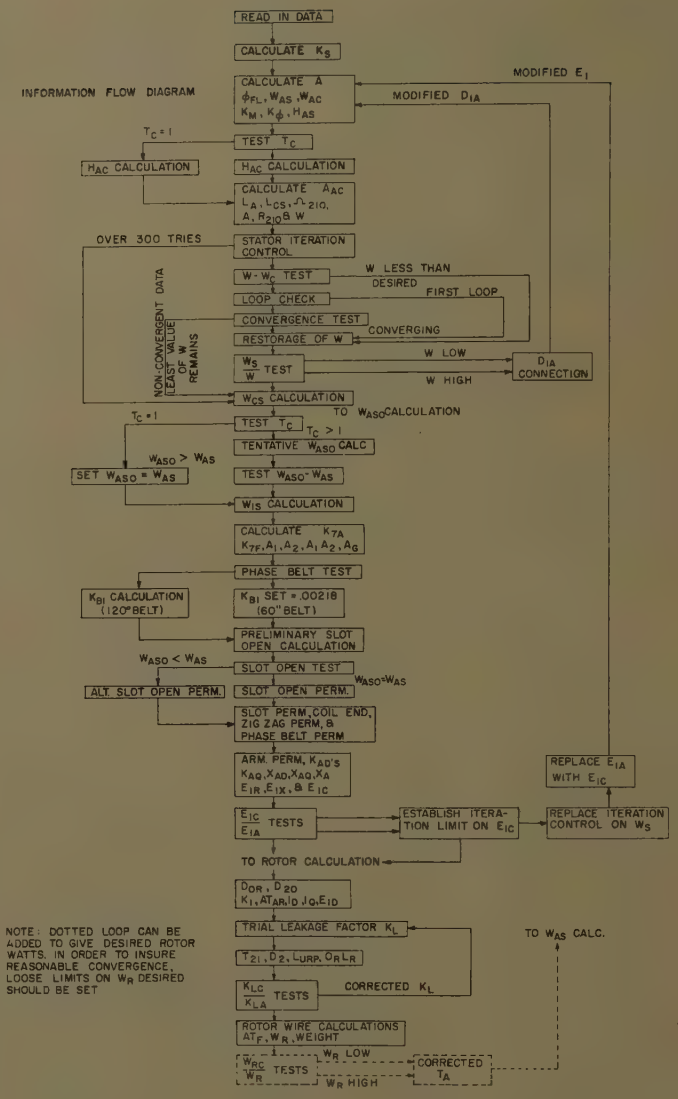


Fig. 7. Generator computer design synthesis deck

Fig. 8. Generator stator

whose impact on weight are rather obscure. Some of these are

1. Diameter-to-length ratio.
2. Pitch and distribution.
3. Air-gap length.
4. Coil throw.
5. Harmonic content.
6. Insulation thickness.
7. Saturation level.
8. Conductor material.

The computer study was used to optimize items 1 through 4.

Program Development

A flow chart of the computer design synthesis deck is given as Fig. 7. The stator is designed first (Fig. 8), requiring stator copper loss to be held to an input figure, (within ±1%). Stator iron design maintains specified flux densities for the given stator turns. Finally, the rotor is designed holding specified flux densities. The correct leakage impedances are fed back in both rotor and stator to give the flux densities correctly.

The beauty of synthesis is that design flux densities are specified as inputs. Therefore, only one point on the magnetization curve need be put into the deck for each design density. The tapering stator tooth is taken care of by an empirical relationship derived to fit calculated magnetization curves for various tip-to-slot-width ratios including slot flux.

One curious effect which was not at first anticipated arose on early trials of the deck. In order to design a stator with the desired stator watts (within ±1%), the stator bore is varied. For example, if stator watts calculated on the first pass are too low, inside diameter is increased. Outside diameter happens to be a piece of input data and so the result is to "squeeze" the slots and so increase the stator watts. Naturally, convergence on the desired watts is assured, because inside diameter is increased until stator watts reach the specified value. However, if stator watts are too high on the first pass, it is possible that on the next pass stator watts will be even higher, even though inside diameter is decreased. This is because as stator bore is decreased, to allow a deeper slot and so lower stator watts, the slot becomes narrower so that a point is reached beyond which slot area becomes *less* rather than greater and so stator watts increase rather than decrease. If this case arises, it simply means that the desired stator watts are too low to be designed into the specified package. To circumvent this situation, the deck compares the newly calculated watts with the previous value for the case of stator watts greater than desired. If the latest value is lower than the previous, calculations continue. If not, the deck accepts the "lowest watts" design and proceeds. The design study was based on the use of rectangular stator slots.

Generator Mechanical Design

Mechanical Warning Device

Recognition and warning of bearing failures in aircraft generators is a safety feature desired by many customers. To accomplish this, a mechanical warning device is incorporated in one tooth of the stator core. This device consists simply of a copper strip embedded in a slot liner which insulates the strip from the tooth iron. Fig. 9 shows the component details. Bearing failure will cause the rotor to pull over onto the stator due to a combination of increased bearing clearance and magnetic unbalanced force. A test on a 15-kva generator showed that with the mechanical warning device in the top tooth (so that gravity would not aid in activating the device), the pullover action of the rotor was ample to smear the copper strip onto the tooth iron and thereby close the circuit to the warning light in the cockpit.

Fig. 9. Generator terminal blocks

Fig. 10. Generator rotor

Quickly Removable Blast Cap

Brush inspection and maintenance are enhanced by incorporation of a simply removable inlet cap. The problem in this case is safety wiring. By using a clamping band design (which springs the band down onto the cap) only one safety wire need be removed in order to release the cap.

Terminal Block Design

Some desirable terminal block design features incorporated in the Electra generator are

1. Barrier to prevent short circuits of cable terminals.
2. Covers to prevent short circuits from extraneous matter.
3. Current conduction from equipment to cable terminals by surface-to-surface contact and not through screw threads.
4. Removal of mechanical loads from cable terminals.

Fig. 11. External view of generator

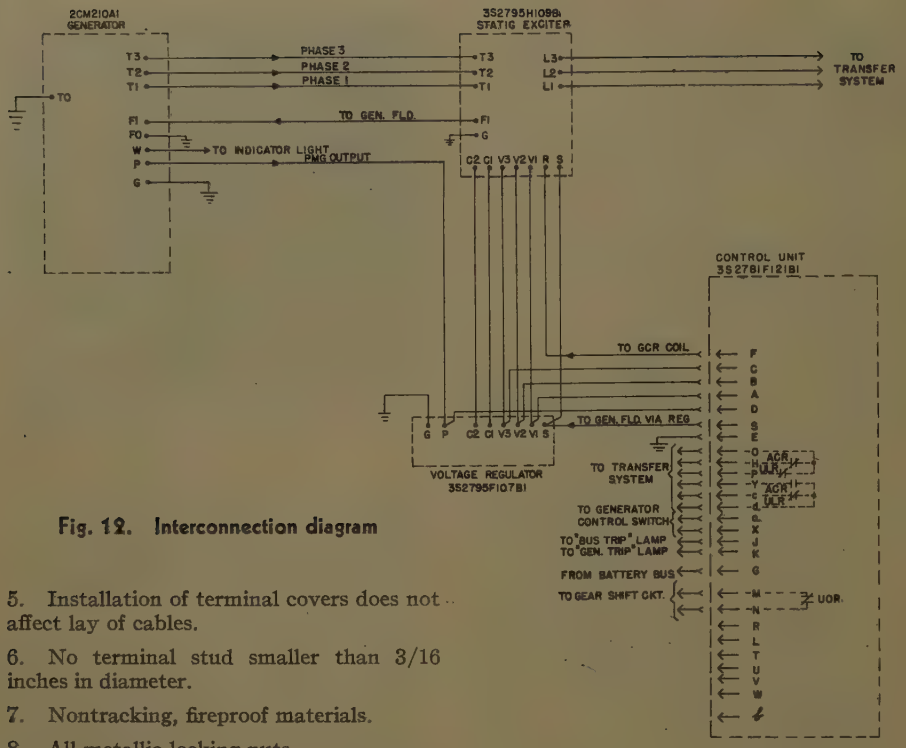


Fig. 12. Interconnection diagram

5. Installation of terminal covers does not affect lay of cables.
6. No terminal stud smaller than 3/16 inches in diameter.
7. Nontracking, fireproof materials.
8. All-metallic locking nuts.
9. Terminal number and color coding.

These features are incorporated into the generator design as shown in Fig. 10. Each cable is separated by barriers from its neighbors. Current is carried from the stator through flag-type connectors against which the cable terminals are placed in direct contact. Further, the main power terminal block cover is designed with an integral cable clamp assembly which binds the main power cables firmly in place as the cover is tightened. This releases the terminal studs and cable terminals of mechanical forces. The auxiliary terminal block has a simple loop clamp installed on the generator directly in front of the block, through which the cables are laid and clamped. Once again mechanical forces are stopped at the clamp.

Fan Cooling

Fig. 11, which shows the rotor assembly, illustrates the wedge fan that is used to provide the cooling required for ground operation of the generator. The wedges are extended on the exhaust end of the generator and provide centrifugal fan action. Typical temperatures with and without this fan are shown in Table II.

Table II. Fan Action CM210 at 6,000 Rpm

Item	With Fan	Without Fan
Rotor temperature rise at full load	143	187
Air rise at full load	55	99
Approximate air flow, cubic feet per minute	200	100

Method of Balance

Fig. 11 also illustrates the end construction of the rotor. High-strength bolts are used to hold down small caps which are placed over the end turns. These caps are made of titanium in the initial construction, and in order to facilitate balance, steel caps of comparable yield and tensile strength are substituted where needed. The weight difference between titanium and steel therefore provides a nice way of obtaining balance.

Physical Characteristics

The use of static excitation has many advantages in the design of the generator. It eliminates the designing of the rotating exciter with its commutator. Because of the elimination of the rotating exciter the cooling problems are reduced. The generator is smaller, lighter, and has less overhung moment. The Electra generator is 11 inches outside diameter and $13^{7}/_{8}$ inches long including blast cap, weighs 91.5 pounds, and has an overhung moment of approximately 435 inch-pounds. The reduced length of the static exciter generator helps to solve the installation problem in the accessory section of the Electra nacelle.

Static Exciter and Regulator

This paper will be limited to discussing those aspects of static exciter and regulator design and performance peculiar to the Lockheed Electra. The interconnection diagram of the system is shown in Fig. 12.

Static Exciter

Briefly stated, the exciter supplies field excitation current by feeding back to the generator, a rectified power proportional

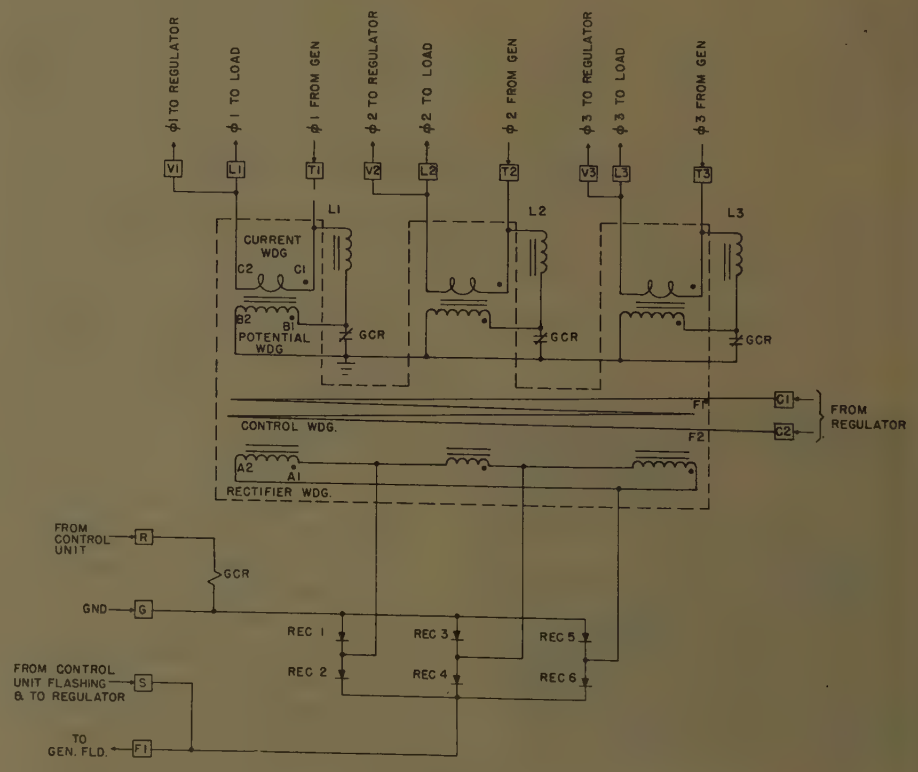


Fig. 13. Static exciter diagram

Fig. 14. Saturable current potential transformer

to a combination of line voltages and currents. The current and voltage feeds are combined vectorially in accordance with the approximate relationship:

$$If = K\mathbf{E}_t + j\mathbf{I}_1 Xd \qquad (1)$$

where

If = field current in d-c amperes
K = constant of proportionality
$\mathbf{E}_t$ = generated voltage vector
$\mathbf{I}_1$ = line current vector
Xd = generator synchronous reactance

Note that $\mathbf{E}_t$ is also a function of Xd.

The voltage and current signals are so proportioned that, for normal loads, the static exciter will supply correct generator excitation to maintain almost constant voltage, without any action from the regulator.

Essentially, the exciter consists of the following parts (see diagram, Fig. 13):

1. Saturable current potential transformer (SCPT), in which the voltage and current feeds are combined vectorially and into which a control signal from the regulator is introduced.

2. Three-phase bridge rectifier which rectifies the exciter output.

3. Linear reactors for shifting the voltage signal in quadrature with the current signal.

4. Generator control relay (GCR) which, when de-energized, short-circuits the SCPT potential windings to de-energize the system. The GCR is controlled by the cockpit generator control switch (GCS) and the control unit as explained in the "Control Unit."

The SCPT is a 3-phase unit employing *C* core magnetic material and is shown in Fig. 14. A continuous regulator control winding is wound around all three phases, the other windings (potential, current, and output) are all wound individually around each core. Use of the *C* cores with their continuous grain orientation makes it possible to use less magnetic material than would be needed for cores made of punched material. A continuous control winding is used to eliminate the problems associated with high voltages induced in individual control windings.

The GCR relay which is used to shut the system down has a relatively easy function by short-circuiting the windings of the exciter. The closing of the GCR contacts kills the excitation to the generator and the system immediately starts building down. Thus, the relay which is used to turn the system off does not have to open any high power conditions. When the static exciter system is shut down, its output voltage is less than one volt and thus, energy is not available to feed a fault. In the rotating exciter systems where the exciter field is opened in order to shut down, a voltage from 50 to 150 volts is available to feed the fault.

Rating

The requirement of 90-kva continuous operation at 5,700 rpm affects the exciter design from the standpoint of heating and output power requirement.

The consideration of maximum temperature in the exciter determines primarily the conductor size of the winding on the SCPT and linear reactor. Tests have been run at 90 kva, 380 cps, and the exciter at +71 C to determine the maximum temperatures of various parts of the exciter. Some of the temperatures reached during a 5-hour run at these conditions were as follows:

SCPT core: +157 C (stabilized after 1½ hours of operation)
SCPT current winding: +170 C (stabilized after 1¼ hours of operation)

The SCPT current winding is a copper strap having a cross-sectional area of 62,500 square mils.

It is interesting to note that the exciter linear reactors reach their maximum temperature while the system is at no load. During tests at +71 C and 380 cps, a reactor core stabilized temperature of +204 C was obtained.

When load current flows through the current coil of the SCPT at high lagging power factors, it induces a voltage in the SCPT winding supplied by the linear reactor. This induced voltage reduces the current through the linear reactor.

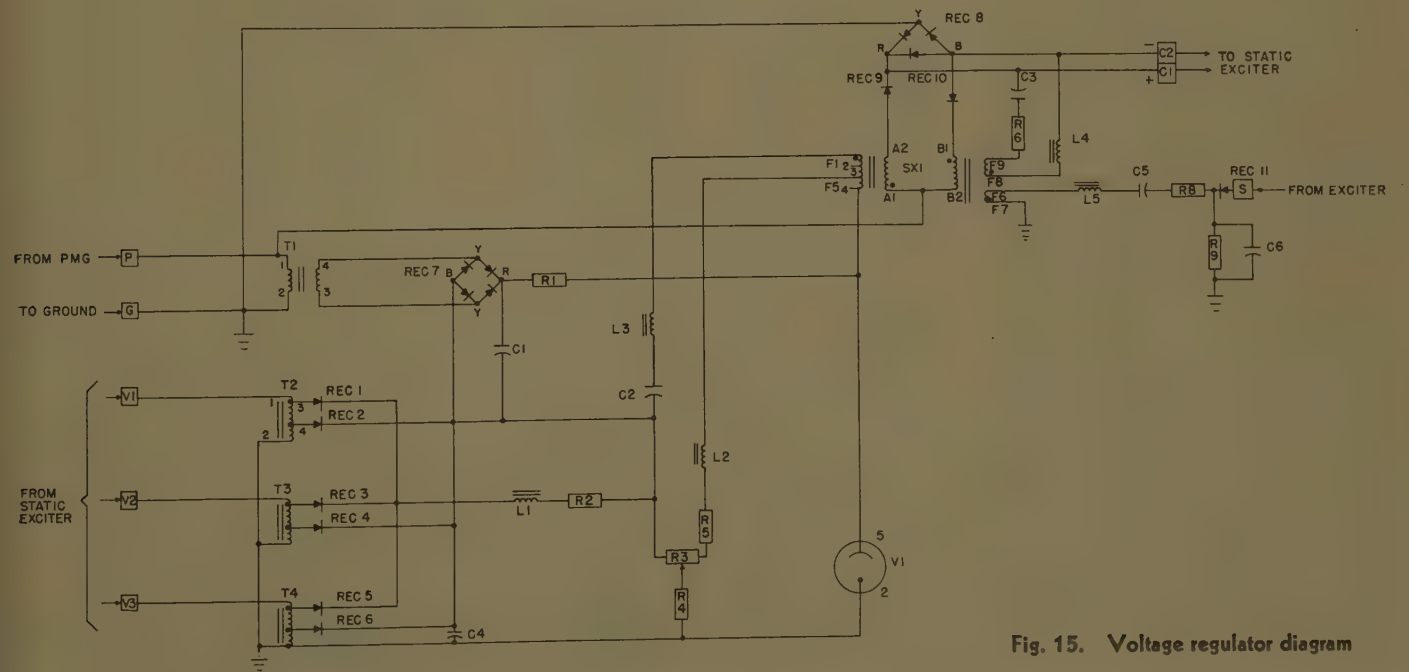


Fig. 15. Voltage regulator diagram

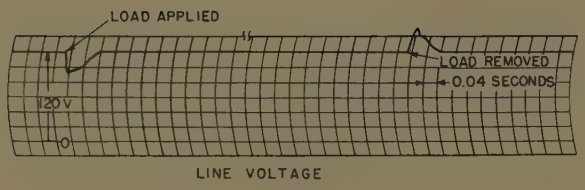


Fig. 16 (above). Transient voltage response

Full load (60 kva at 0.75 pf) applied and removed frequency = 380 cps

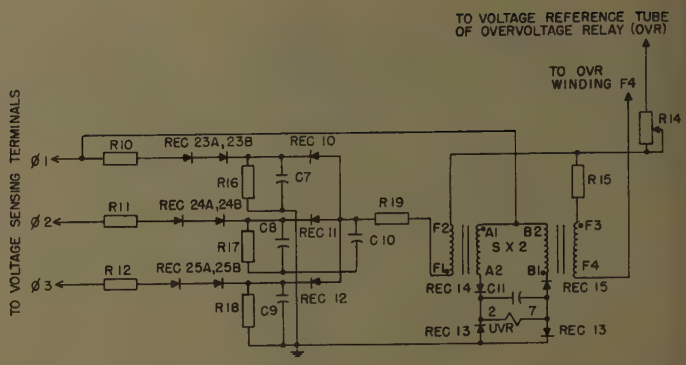


Fig. 17 (right). Undervoltage relay diagram

The static exciter is a highly efficient means of conversion. Following is a tabulation of the exciter losses and efficiency while operating at 1.5 loads. These figures are based on 380-cps operation with generator field resistance = 1.2 ohms, 50-ampere field current, exciter components at their maximum temperature with +71 C ambient, and an assumed 2-volt drop across the rectifiers. Loss figures are probably good to ±10% but the efficiency value is more accurate.

I^2R losses:	
Linear reactor =	16.8 watts
Output winding =	62.9 watts
Potential winding =	6.0 watts
Current winding =	32.0 watts
Control winding =	4.3 watts
Total I^2R losses =	122.0 watts
Core loss (SCPT) =	60.0 watts
Core loss (Linear Reactor) =	10.0 watts
Rectifier loss =	100.0 watts
Total losses	292.0 watts
Efficiency = 92.2%	

The matter of maximum design field voltage is a joint exciter-generator consideration. Essentially, the exciter and regulator combination form a constant current device. For any given generator load and frequency the regulator will attempt to maintain constant field current as the field resistance varies, by varying the control current, which in turn causes the field voltage to vary. With no regulator output, the exciter will put out a maximum field voltage. This value determines the maximum excitation current that can be delivered to the generator for any value of generator field resistance.

Voltage Regulator Electrical Design

The regulator consists of a sensing circuit with input rectifiers and a glow tube reference, a single-stage magnetic amplifier, a stabilizing circuit, and a transformer-rectifier for converting permanent magnet generator (PMG) power to direct current. See Fig. 15.

The glow tube reference was chosen for its lightweight, rugged construction, and accurate performance over a wide temperature range. Service experience on thousands of static regulators testifies as to its reliability.

Highest phase voltage sensing is a circuit design that eliminates a dangerous condition that can affect safety in flight. In most systems in use today, the voltage regulator senses the average of the three line voltages and regulates on this average. It is possible (and has occurred) for one or two phases to have an overvoltage with reduced voltage on the other phases. A faulted feeder or line in the distribution circuit can cause this unsafe condition with 3-phase average sensing. To see the reason for this, consider this example. *L1* is faulted at the bus and at zero voltage and *L2* and *L3* are still feeding power. The average voltage is low and the regulator increases excitation. Since overvoltage protection sees the same value as the regulator, it does not trip. The voltage on *L2* and *L3* can rise as high as 160 volts and cause damage to essential loads so that they would not be available even if transferred to another generator.

A solution to this problem is regulation of the highest phase voltage. This is accomplished by the addition of a capacitor in the regulator sensing circuit. However, a straight highest phase sensing has a disadvantage under unbalanced loading in normal operation. The resulting unbalance in voltage causes the lowest phase voltage to decrease below normal by the amount of the spread between the lowest and highest phase voltage. This can cause it to become undesirably low. With average sensing, the unbalance would cause the highest phase voltage to rise somewhat, and the lowest phase to drop less. The total spread would not be decreased noticeably, as it is a function of the generator characteristics and line drop, but the departure from normal of the lowest phase voltage would be less.

The regulator uses a sensing circuit which has the advantages of both types of sensing. It uses two taps on an autotransformer on each phase. The higher taps feed through rectifiers, a reactor, and a resistor, so that essentially their smoothed-out average value is impressed on a capacitor. The lower taps feed through rectifiers directly to the capacitor with peak values normally just under the voltage from the reactor and resistor. Thus, in normal operation the average phase voltage on the capacitor constitutes the signal to the reference circuit. However, if unbalance causes one phase voltage to rise enough so that its peak value from the lower tap exceeds the average value from the higher taps, this highest phase voltage will become the signal to the reference circuit. Thus, average sensing is used for normal operation, but highest phase sensing "takes over" when unbalance causes one phase voltage to rise to the limiting level.

The error signal from the reference circuit is amplified through the magnetic amplifier whose output in turn controls the saturable current-potential transformer in the exciter. Close voltage regulation is provided by the use of a high-gain magnetic amplifier in the voltage regulator. The system servo loop includes the time constants of the generator, exciter, and regulator. In order to provide stability to this loop and yet maintain fast transient voltage response to load changes three stabilizing circuits are included in the regulator as follows:

1. Lead network.

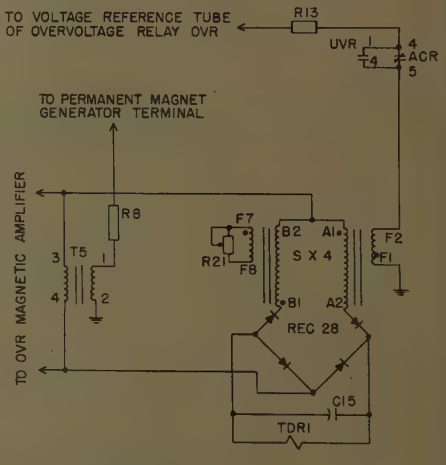


Fig. 18. Time delay relay 1 diagram

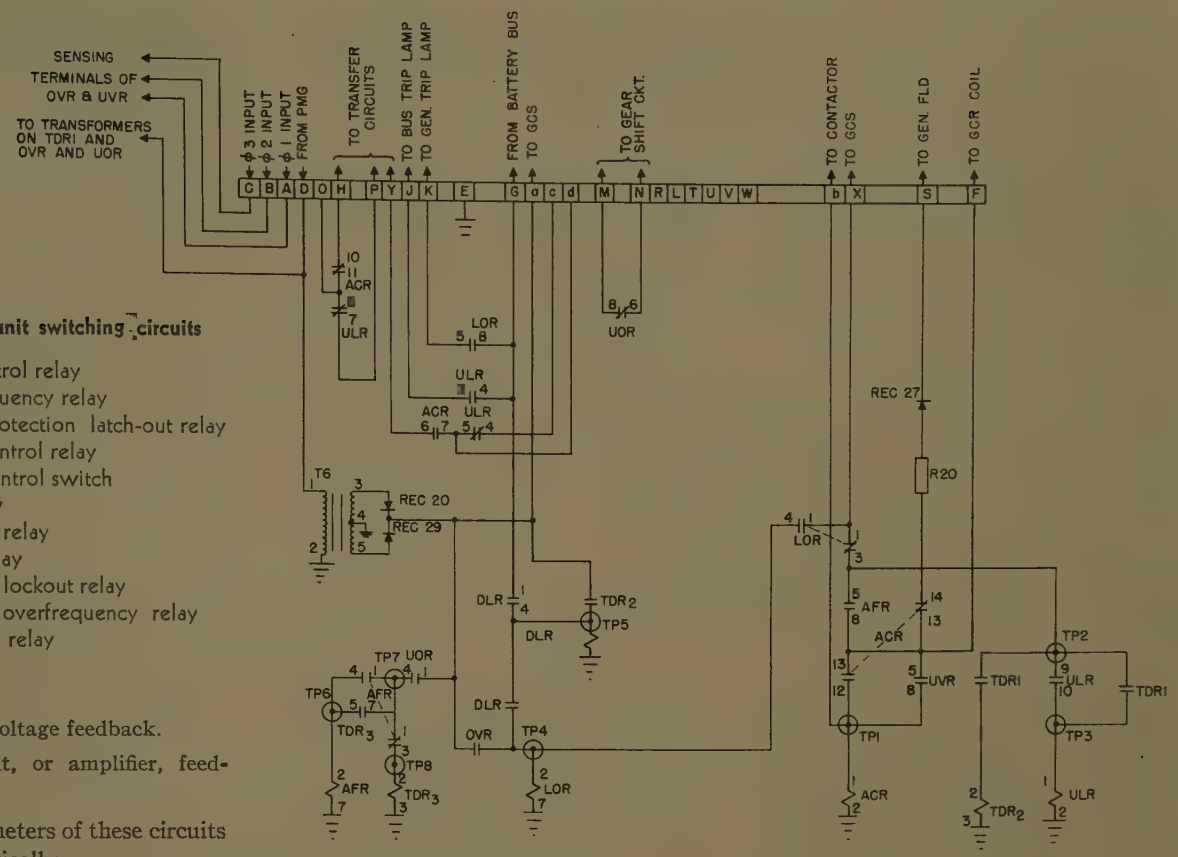


Fig. 19. Control unit switching circuits

ACR = auxiliary control relay
AFR = auxiliary frequency relay
DLR = differential protection latch-out relay
GCR = generator control relay
GCS = generator control switch
LOR = lockout relay
OVR = overvoltage relay
TDR = time delay relay
ULR = undervoltage lockout relay
UOR = under and overfrequency relay
UVR = undervoltage relay
TP = test point

2. Generator field voltage feedback.

3. Regulator output, or amplifier, feedback.

Values for the parameters of these circuits were obtained empirically.

A typical transient response trace showing full load at 0.75 power factor applied and removed at 380 cps is seen in Fig. 16.

Tests over the temperature range −55 C to +71 C indicate that the regulation up to 90 kva is well within the specified limits. Typical data taken with a set of prototype equipment are shown in Table III. These data were taken with a no-load line-to-neutral voltage setting of 116 volts at room temperature and the tests were run at 380 cps.

In the case of the +71 C test, the equipment had been operating at 90 kva continuously during the period of the test. For the other tests, the readings were taken with the equipment operating for approximately 10 minutes at full load.

Control Unit

Undervoltage Protection

The undervoltage protective circuitry must distinguish between two types of sustained undervoltages; those caused by excessive or fault current drawn by the load, and those caused by a fault or malfunction within the generator system. In each case, it must take proper action in order to prevent damage to the aircraft or its equipment. At the same time, it must be capable of ignoring transient undervoltages caused by sudden load application. Conventional differential protection is included in the Electra production models but was not included in the prototype systems.

Table III. Typical Data From Temperature Tests

Temperature Conditions	Average 3-Phase Voltage Readings		
	No Load	60 Kva	90 Kva
Room temperature	116.0	117.0	116.8
After 4 hours at +71 C	117.5	118.2	115.0
After 2 hours at −55 C at no load	117	118.0	117.8

The undervoltage magnetic amplifier circuit shown in Fig. 17 is used to operate the undervoltage relay (UVR) which controls the protective circuitry. The UVR is de-energized when the voltage on one or more phases is below a predetermined minimum and energized when all the phase voltages are above this minimum. The trip point, tentatively set at 105 volts ±2%, is determined by the values of *R10* through *R18* and *R15*; an adjustable resistor *R14* is used to provide a fine adjustment of the trip point and compensate for variations in reference tubes, relay trip points, etc.

Consider now the functional sequence of events in the case where, with all three phases at normal voltage, the voltage on one or more goes below the trip point:

1. If the voltage returns to normal within 3.0 seconds±25% (by removal of excessive load from the bus or tripping of load breakers), the UVR is re-energized and there is no action taken by the protective circuitry.

2. If the voltage does not return to normal within the 3 seconds, the bus will be disconnected from the generator.

3. If the voltage then returns to normal within an additional 1.0 second±50% the generator remains energized but the bus will be locked off the system.

4. If the voltage does not return to normal within the additional 1 second, indicating a feeder or generator fault, the generator will become de-energized and lock out of the system. The bus, however, will be transferred to another generator.

In order to obtain the first time delay, a time delay magnetic amplifier (*TDR1*) circuit is employed as shown in Fig. 18. The amplifier is powered from the PMG via a saturable transformer which provides a constant volt-second-per-cycle source of energy in order to make the time delay independent of PMG output fluctuations. Note that the saturable transformer output varies directly with frequency but over the usable frequency range 380–420 cps, this variation has little effect on the time delay.

Construction of the *TDR1* magnetic amplifier is such that it will keep the *TDR1* relay de-energized as long as current flows in the bias winding *F1-F2*. Removal of the current will start the magnetic amplifier "timing out" and after the delay determined by short-circuited winding *F6-F7* and resistor *R21* setting,

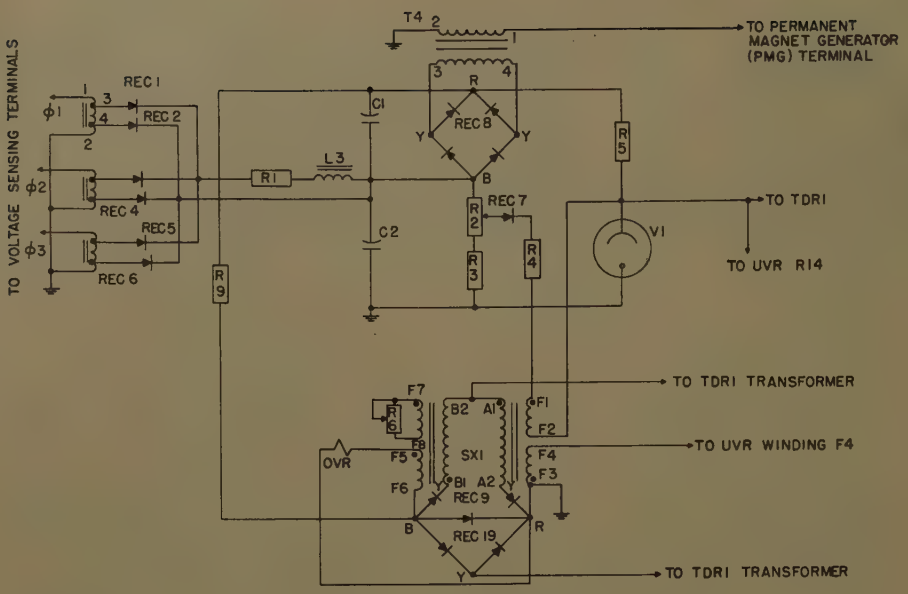


Fig. 20. Overvoltage relay

the *TDR1* relay will become energized. Current for the bias winding is obtained from the voltage reference tube *V1* and is held to approximately 3.5 ma by limiting resistor *R13*. In series with the bias winding are a set of normally open UVR contacts and normally closed auxiliary control relay (ACR) contacts. The ACR is energized when the UVR picks up after field flashing and remains energized regardless of UVR operation, as long as the generator control switch (GCS) is closed and the frequency is within prescribed limits. This contact arrangement prevents removal of bias current and timing out when the generator is up to speed but the field has not been flashed, yet allows the relay to be energized when the voltage falls below the trip point after build up.

With reference to Fig. 19 which shows the unit's internal switching circuits, the sequence of operations after the dropping out of the UVR is as follows: The *TDR1* magnetic amplifier starts timing out. If the voltage returns to normal within the 3 seconds period, the UVR will be re-energized and the timing out action will be halted. If the voltage does not go back to normal, the *TDR1* relay will be energized. A set of *TDR1* contacts will then energize the undervoltage lockout relay (ULR), contacts of which operate the transfer circuitry to take the generator off the bus. The ULR is self-sealing in order to prevent recycling of the circuitry if the voltage returns to normal. In order to drop out the ULR, it is necessary to open the GCS and then close it.

If now the voltage returns to normal, there is no further operation of the circuitry. The GCS may be opened and then closed to reset the ULR, provided the cause of the undervoltage has been eliminated. If, however, the voltage does not return to normal, a set of normally open *TDR1* contacts will apply power to the coil of time delay relay *2* (*TDR2*), a thermal-type time-delay relay. In one additional second the *TDR2* contacts will close, energizing the differential lockout relay (DLR). A set of DLR contacts will energize the lockout relay (LOR) which will interrupt the coil current to the generator control relay in the static exciter and the bus transfer is reset. De-energization of this last relay will disable the exciter, thereby de-energizing the generator. Both the DLR and LOR have self-locking contacts, the former fed from the battery bus and the latter from the unit's TR via the GCS. Thus, to attempt to re-energize the generator it is necessary to open and close both the GCS and the connection between the unit and the battery bus.

Overvoltage Protection

Sustained overvoltages can arise from several causes, including the following:

1. An internal failure of the voltage regulator.
2. Opening of the leads between the regulator and exciter.

Transient overvoltages arise from field flashing, and removal of loads and faults. In order to eliminate nuisance tripping and yet obtain fast operation of the protective circuitry on sustained overvoltages, a magnetic amplifier having a constant volt-second trip time delay is employed to operate a relay within the protective circuitry; see Fig. 20. The overvoltage relay uses highest phase takeover sensing of voltage. Expressed mathematically:

$$K = \int_0^T (E_T - UTP)dt$$

where

K = overvoltage relay trip time constant in volt-seconds
UTP = ultimate trip point in volts rms, or voltage which will cause OVR to trip after being applied for an infinite time.
E_T = the transient rms voltage. Only the region $E_T > UTP$ is of interest.
T = time elapsed in seconds between the application of E_T and the energizing of the OVR relay.

Values of UTP and K will vary somewhat depending upon whether one, two, or three phases are high. If a constant value of E_T is considered, then the settings for the Electra are

$(E_T - UTP) = 10$ volts
$K = 4.5$ volt-seconds

with three phases high. These values represent the minimum that can be tolerated without risking nuisance tripping.

Current for the bias winding *F3-F4* is obtained from the voltage reference tube *V1* through current-limiting resistor *R15* (in actual practice the OVR bias winding is in series with the UVR bias winding and the bias current is obtained through *R14* and *R15*). The ultimate trip point is set by *R2* and the time constant by *R6*. The OVR relay is de-energized at normal voltage and energized by overvoltages.

By referring to Fig. 19, it will be seen that the overvoltage protection is accomplished in a manner similar to under-

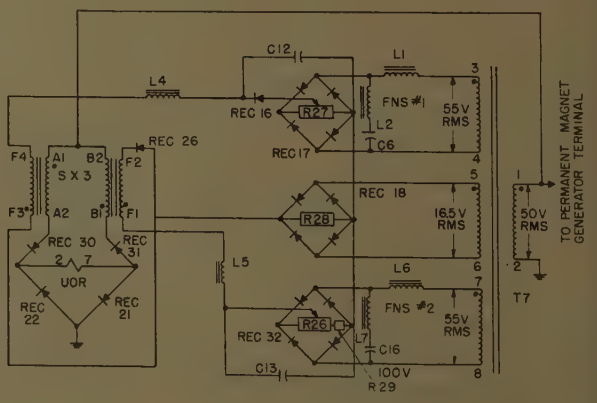


Fig. 21. Under- and over-frequency relay diagram

voltage protection, in that a set of OVR contacts is used to energize the LOR, which in turn de-energizes the generator. Opening and then closing of the GCS will reset the protective circuitry and re-energize the generator. However, if the overvoltage fault remains, the overvoltage relay will again trip and de-energize the system.

Under- and Overfrequency Relay

Engine speed in flight is normally constant at 13,800 rpm. However, for ground idle the engine speed is reduced to 10,000 rpm. Therefore, in order to supply essentially constant frequency power for all conditions, a 2-speed gear-shifting device is incorporated into the airplane. Co-ordination with the gear-shifting operation requires the incorporation of the time delay *TDR3*.

Fig. 21 shows the frequency-sensing circuit used to control the under- and overfrequency protection circuitry. Essentially the circuit consists of a magnetic amplifier operated under- and overfrequency relay (UOR), two frequency-sensitive networks (FSN no. *1* and FSN no. *2*) and a reference network. The difference between the FSN's and the reference network outputs are fed through blocking rectifiers *R16* and *R26* to the two control windings of the magnetic amplifier. When no current flows in either control winding, which will be the case when the frequency is within the 360–425-cps band, the UOR relay will be energized. Outside this band, current will flow in one or the other winding and the relay will be de-energized. Frequency values at which current will start flowing in the control windings are determined by the settings of *R26* and *R27*.

Two FSN's are used because of the relatively wide band of frequencies over which the system must operate. If only one FSN were used, the curve of E_{out}/E_{in} versus frequency would have to be relatively flat to include both extremes. This would result effectively in a low milliamperes-per-cycle gradient for the magnetic amplifier control winding with a consequent large differential between relay pickup and dropout. With two networks, each one can be designed separately to give a high gradient in the relay trip region. Frequency response curves for both networks are shown in Fig. 22.

Consider now the operation of the UOR as frequency is varied from below 360 cps to above 425 cps. Starting below 360 cps, the output of FSN *1* and *2* will be greater than the output of the reference network. Hence current will flow through blocking rectifier *REC16* and winding *F3-F4*. UOR relay will be de-energized. No current will flow in *F1-F2* because of the position of *REC26*. Above 360 cps, the output of FSN *1* will be less than the output of the reference network and no current will flow in *F3-F4*. The output of the reference network will still be less than that of FSN *2* so no current will flow in *F1-F2*. The UOR relay will now be energized. Above 425 cps, the output of the reference network will be greater than that of FSN *2* and current will flow through blocking rectifier *REC26* and winding *F1-F2*, thus de-energizing the UOR relay.

The protective circuitry is shown in Fig. 19. It will be noted that a set of normally open UOR contacts and a set of normally closed auxiliary frequency relay (AFR) contacts are in series with *TDR3* coil. The *TDR3* relay is of the thermal timer type. Assuming that the UOR relay is initially de-energized, both the *TDR3* and AFR relays are de-energized. The AFR contacts in series with the output to the GCR coil will be open and the generator will be de-energized. Energizing the UOR will apply current from the PMG TR to the *TDR3* coil. After the 3 seconds' delay the *TDR3* contacts in series with the AFR coil will close. The AFR will be energized and lock itself in through its normally open contacts, at the same time, it will remove current from the *TDR3* coil allowing it to cool off and reset. The AFR contacts in series with the GCR coil will close and (assuming the GCR is closed) allow the field to be flashed.

Field Flashing

Since the generating system is completely self-excited, a "flashing" source must be provided for initial generator buildup. The source of the flashing power is the PMG. The flashing circuitry is shown in Fig. 19.

Alternating voltage from the PMG is rectified by transformer *T6* and rectifiers *REC20* and *REC29* which produce a full-wave rectified signal of approximately 28 volts d-c at 6,000 rpm. Assuming that the generator is up to speed, the LOR is not energized, and the GCR is closed. current from the transformer-rectifier will flow through the GCR, normally closed LOR contacts and normally open AFR contacts to the GCR coil. This will lift the short circuits from the exciter SCPT potential windings and permit the exciter to function. Current for flashing will be fed to the generator field through current limiting resistor *R20* and the normally closed ACR contacts. As soon as the line voltage rises high enough to energize the UVR, the ACR will be energized and lock itself in thus removing the flashing power from the field. Blocking rectifier *REC27* is used to prevent the possibility of the field voltage feeding back to the GCR when the LOR is energized and the ACR is de-energized.

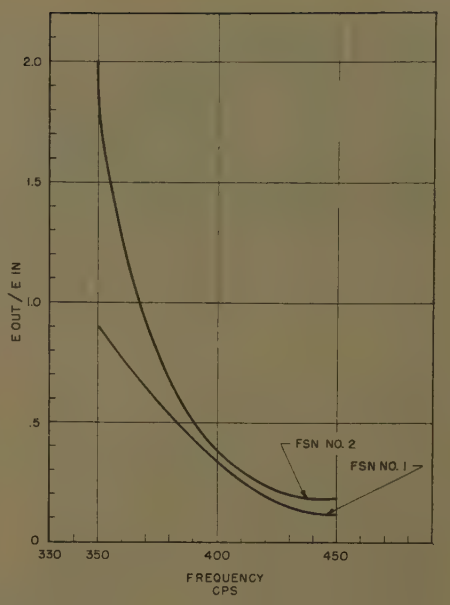


Fig. 22. Response of under- and overfrequency relay

Product Design of Electra Controls

It is not merely outstanding electrical performance that makes a unit acceptable for use in a modern aircraft, but reliability light weight, and maintainability as well. A major factor contributing to successful equipment embodying these characteristics is the simultaneous consideration of the problem of both installation and product design thereby providing the best over-all compromise. Major components within a unit usually determine its optimum shape and size. The shape, however, may or may not provide the optimum over-all result and, therefore, should be considered together. In addition, a willingness on the part of both equipment manufacturer and user to make modifications as required to acquire this goal is a must. The design of the equipment for this system is an example of this type of co-operation.

The static exciter weighing $24^1/_2$ pounds presented a major problem in face of the requirement that no shock mounts be used in meeting the specification vibration environment of 10 *g* up to 250 cps. The philosophy of design of a unit of this weight and size was to utilize the relatively stiff construction of the oblong-shaped SCPT as a backbone from which the few remaining com-

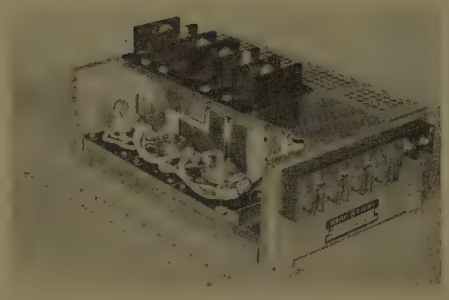


Fig. 23. Static exciter

ponents could be supported; see Fig. 23. This is in contrast to the usual scheme of building a chassis onto which the components are mounted. The SCPT utilizes *C*-type cores banded securely into a U-shaped member which runs from front to back. The terminal board is mounted on the top of this member and with the connections to the transformer make an integrated assembly. Because of its location in the main bus of the system, maximum protection from line faults within the exciter is provided. This is accomplished by using a rigid copper bus assembly flat against the U-member and sandwiched tightly between flat sections of silicone-glass laminate which insulate it from ground and against faults caused by foreign objects falling against it. Full-height barriers for all terminal boards are provided for the same reason.

The six silicon rectifiers are mounted onto flat radiators which in turn are mounted horizontally along the side of the transformers assembly adjacent and below the bus bars. Thus, with the perforated cover removed, all rectifiers and the bus assembly are readily accessible. The three linear reactors are mounted on the same member as the transformers, adjacent to one another across the front. The control and field terminal boards are mounted on the front. The main terminal boards on the exciter are made such that the cables to the transfer box can be connected from either right or left, thereby allowing identical units to be used for both sides.

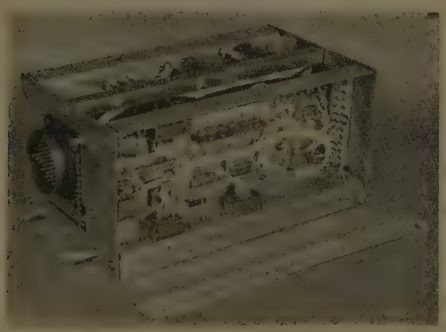


Fig. 24. Control unit

An example of the afore-mentioned modifications to meet the desired over-all objective was a change at the prototype stage from an exciter having the main terminals at the front and ATR-type mounts to the foregoing design. The ATR mounts with pins in the rear and clips in front were abandoned as being inadequate for holding down an unshock-mounted unit. They imposed an undue weight penalty as well in this application.

The control unit is much more conventional in design; see Fig. 24. This unit which resembles electronic rather than power equipment, also must survive in a 10 *g* up to 250-cycle environment without shock mounts.

The best technique for accomplishing this is the putting of complete functions into sturdy modules. This was considered but not used because of the economies of replacement; a $25 to $50 limit on unrepairable subassemblies being imposed by the user airlines.

Printed circuits were also considered and a prototype was built which used two 8-inch by 5-inch boards; however, this, too, was abandoned because the large number of bulky components such as relays, transformers, magnetic amplifiers, etc., made a dense package difficult. The problem of damage to the print circuit during component replacement was also a factor in the decision.

The construction used involved three full-length pan-type chassis, one on each side and one at the bottom. These combined with a U-shaped top cover and a flat bottom cover comprise rigid box members which withstand the vibration. Each chassis contains one or more complete functions such as an overvoltage relay assembly which expedites trouble-shooting or replacement. Solder terminal boards are provided on each panel for this purpose. Plugs, while of greater convenience, are not used because of their lesser reliability.

Components are mounted on the pan chassis in the conventional manner but with the wiring portion facing the outside where it is completely accessible when the covers are removed. Each chassis can be swung out at right angles to the panels main axis for access to both sides since the wiring harness feeds to a single terminal board on each end. Important test points are brought out to their same terminal board.

Over-all dimensions of the control unit are 6 by 7 by $11^3/_4$ inches and it weighs 13 pounds; see Fig. 25.

The small size of the regulator (6 by 6 by $5^1/_2$ inches and $5^1/_2$ pounds) allows use of a still more conventional pan-type

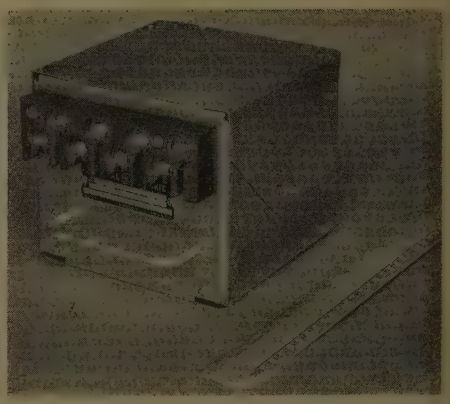


Fig. 25. Voltage regulator

chassis although it is made L-shaped to give greater mounting area and rigidity. The terminal board is the same as used on the front of the exciter.

Appendix. Solution for Optimum X_d

$$X_d^2(X_d+a)=b$$

Since

$$\frac{b^2}{4}+\frac{a^3}{27}>0$$

(a is always +) then:

$$X_d=\pm 2\sqrt{\frac{a}{3}}\cot 2\phi$$

where

$$\tan\phi=(\tan\psi)^{1/3}$$

$$\cot 2\psi=\sqrt{\left(\frac{27}{4}\right)\left(\frac{b^2}{A^3}\right)}$$

For small values of ψ (i.e. small a),

$$\tan 2\psi\rightarrow 2\psi$$

and so

$$2\psi=\sqrt{\frac{4a^3}{27b^2}} \qquad \psi=\frac{1}{b}\sqrt{\left(\frac{3}{a}\right)^3}$$

Then by trigonometric identity:

$$\cot 2\phi=\frac{1-\tan^2\phi}{2\tan\phi}=\frac{1-(\tan\psi)^{2/3}}{2(\tan\psi)^{1/3}}\rightarrow$$

$$\frac{1-\psi^{2/3}}{2\psi^{1/3}}=\frac{1-\frac{1}{b^{2/3}}\left(\frac{a}{3}\right)}{2\left(\frac{1}{b}\right)^{1/3}\sqrt{\frac{a}{3}}}$$

So

$$X_d\rightarrow(b)^{1/3}\left(1-\frac{a}{3b^{2/3}}\right)$$

Reference

1. Some Applications of Magnetic Amplifiers in Aircraft Generator Protective Systems, D. L. Plette, J. W. Butler. *AIEE Transactions*, vol. 76, pt. II, 1957 (*Paper no. 57-481*).

Salient-Pole Permanent-Magnet Alternators for High-Speed Drive

ALLEN T. PUDER
NONMEMBER AIEE

FRITZ STRAUSS
MEMBER AIEE

MODERN prime movers, especially turbines, and new applications such as missile auxiliary-power units, have established requirements for reliable, efficient, high-speed machines for the generation of electric energy. The specific feature of the permanent-magnet alternator, i.e., elimination of the wound field and, hence, external excitation and slip rings, makes this type of machine superior to conventional machines for these applications. Reliability is enhanced by the elimination of arcing contacts which cause radio interference, brush operating difficulties at high altitudes, and explosion danger in confined, fuel-laden atmospheres.

This paper is concerned with a description and the design problems of some of the more recent units which have been built and tested in production. Rated outputs up to 10 kva have been realized at speeds to 24,000 rpm. Greater outputs and higher speeds have been achieved in experimental machines.

Nomenclature

B_g = air-gap flux density
B_{gm} = air-gap flux density reduced to magnet area
B_{sc} = air-gap flux density at short-circuited stator winding
B_{st} = flux density of magnet at point of stabilization
C_m = armature reaction factor[1]
C_p = ratio of average to maximum of the field form
C_1 = ratio of the fundamental to the actual maximum field form[1]
D = stator outside diameter
d_r = rotor diameter
E_l = alternator line voltage
E = alternator electromotive force (emf) per phase
E_{sc} = alternator emf at short circuit
F_{dm} = armature reaction, ampere-turns per pole
g_e = effective air gap
H_{dsc} = demagnetizing force of stator field (ampere-turns per inch magnet length per pole) at short circuit
H_{st} = demagnetizing force in the magnet (ampere-turns per inch magnet length) at the point of stabilization
h = width of stator core
I_{sc} = stator current per phase at short circuit
k = design constant
k_d = distribution windings factor
k_1, k_2 = proportionality factors
l = axial length of stator iron
m = number of phases
n_e = total number of effective stator conductors
P_g = air-gap permeance
P_{ge} = combined permeance of air gap and stator core
P_l = in-stator leakage permeance[2]
P_m = spatial permeance of permanent magnet[2]
P_o = out-stator permeance[2]
p = number of poles
r = resistance of stator winding per phase
x_l = stator leakage reactance per phase
z_{sc} = stator short-circuit impedance per phase
β = angle of stabilizing shearing line
γ = angle of working shearing line
δ = angle of "in-stator" shearing line
μ_e = effective permeability
μ_{st} = slope of stabilizing shearing line
μ_p = permanent permeability, slope of minor hysteresis loop[2]
σ = stacking factor of stator laminations
τ = ratio of reluctances of back iron and air gap

Mechanical Design

Basic design objectives for developing the permanent-magnet alternator for high-speed applications are

1. Maximum reliability under extreme storage and operating environmental conditions.

2. Minimum weight and space for a maximum output.

With these objectives in mind, alternators have been designed to couple directly to high-speed prime movers. This arrangement eliminates the requirement for any complex power transmitting components and minimizes the number of supporting bearings required.

These alternators are radial air-gap machines, of compact design and high efficiency. The stationary armatures of the machines are of conventional type, with single- or multiple-phase windings. The rotating fields utilize high-energy *Alnico V* permanent magnets. Fig. 1 shows a typical rotor construction, designed in a basic cage-type structure, which is die-cast with an aluminum alloy. The cage structure is built up of stainless-steel (nonmagnetic) plates, with rivets holding the laminated pole tips, *Alnico V* permanent magnets, and an inner magnet tie block in place around the shaft.[3] In a 2-pole rotor structure, the inner magnet tie block and through shaft are eliminated and flanged drive hubs are used, as shown in Fig. 2.

An interference fit is used in the cage assembly to support the magnet and to maintain a tight magnetic circuit. The magnet is used as a structural member in compression only, since *Alnico V*, with its exceptionally good magnetic properties, is a hard and brittle material with low tensile strength.

Die casting of the preheated rotor assembly is performed using a special casting die into which molten aluminum is forced from one end of the die to the other. High die-casting pressures ensure filling the mold and all internal voids, including the volume between the magnets and the damper-bar openings, with a dense material. Aluminum is used, since it is readily cast within the required temperature range, has good mechanical strength properties, and is a good electric conductor. Being an electric conductor, the aluminum serves as a closed turn, surrounding each magnet for the protection of the permanent-magnet field under transient short-circuit conditions. It also improves the waveform through the damper-bar winding and, in the case of an unbalanced load, reduces the unbalance of the phase voltages.

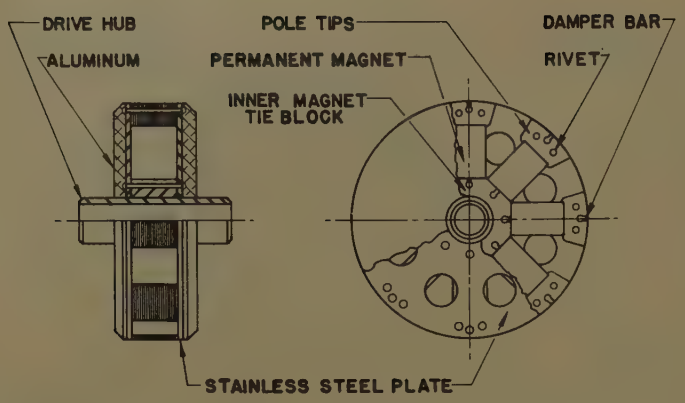
All critical parts used in the rotor structure are made from dies, in a punch-press operation, for economy in production quantities. The *Alnico V* magnets are cast in blocks, to provide maximum magnetic properties, and then ground to the finished dimensions. With all parts being identical for a given rotor diameter and number of poles, the stack height can be adjusted to meet a wide range of output ratings, using a minimum stock of parts and special dies.

High-Frequency Alternator Design

The output frequency of permanent-magnet alternators, operating at speeds up to 24,000 rpm and constructed as described, has been limited in a small rotor by the impossiblity of magnetizing many closely spaced poles of opposite polarity. This limitation is eliminated by the unique rotor construction shown in Fig. 3. By alternating the permanent

Paper 57-489, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE East Central and Middle Eastern District Meeting and Air Transportation Conference, Dayton, Ohio, May 7–9, 1957. Manuscript submitted January 15, 1957; made available for printing May 17, 1957.

ALLEN T. PUDER is with the Ruckstell Corporation, Azusa, Calif., and FRITZ STRAUSS is with the Christie Electric Corporation, Los Angeles, Calif.



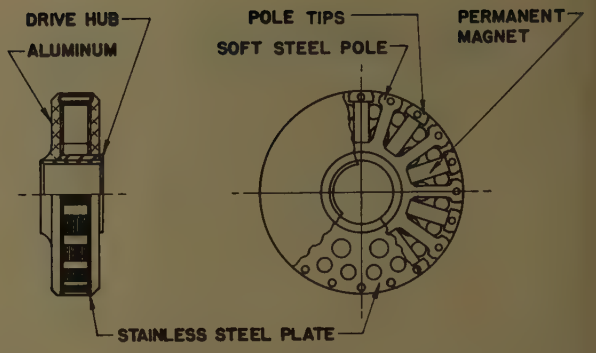


Fig. 3. Twenty-pole rotor construction with alternate permanent magnet and soft-steel poles

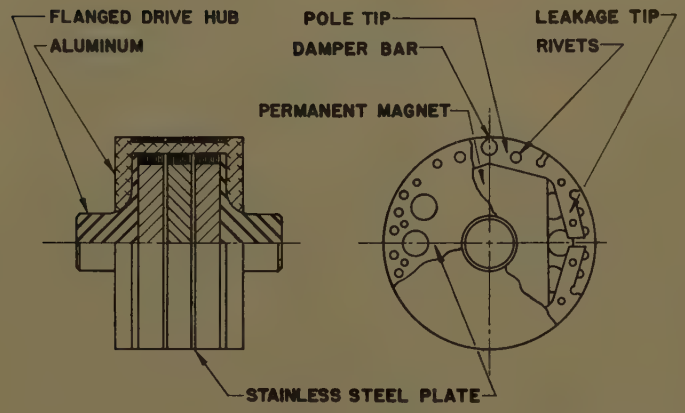


Fig. 1 (above, left). Eight-pole rotor, sectioned to show parts of typical cage structure with hollow shaft

Fig. 2 (left). Two-pole rotor, sectioned to show flanged drive hub and other parts of cage structure

magnet and soft-steel poles around the rotor circumference, so that the permanent magnets are of the same polarity, the entire rotor can be magnetized by a unidirectional radial field.[4] The soft-steel poles are saturated during the magnetizing impulse and immediately change to their proper polarity, becoming poles of a normal fully magnetized rotor. A typical small rotor of this construction, with its magnetizer, is shown in Fig. 4.

Stabilization

The permanent-magnet field of the alternator must be stabilized in order to avoid any voltage change subsequent to the application of a specified load condition. Stabilization methods commonly used, and used in the salient-pole alternator, are: 1. load stabilization, 2. short-circuit stabilization, and 3. air stabilization. The method or degree of stabilization used depends upon the requirements of a particular application. A brief description of each method follows.

Load Stabilization

Load stabilization is used in applications where weight is of utmost importance. Minimum weight is achieved by retaining the maximum generating flux. To accomplish this the rotor is assembled by means of a special magnetic keeper and then stabilized by applying the specified maximum load. Loads exceeding this stabilizing load will continue to decrease the operating voltage until a short-circuit condition is reached, beyond which no further demagnetization can take place.

Short-Circuit Stabilization

Where a short circuit may accidentally be applied to the alternator, short-circuit stabilization is used. The rotor is assembled as described in the foregoing under "Load Stabilization." A sudden short circuit is then applied to the armature winding, while the alternator is operating at rated speed. Since a sudden short circuit causes the greatest possible load demagnetization, there will be no further decrease in the operating voltage, unless the alternator is disassembled.

Design Calculations

The point of stabilization on the demagnetization curve of the permanent magnet, as well as the air-gap density, can be obtained by consideration of the following conditions:

1. The magnetomotive forces of rotor and stator must be in balance.
2. The sum of the voltages in the short-circuited stator winding must be zero.

As shown in the appendix, the air-gap flux density at no load, reduced to the magnet cross section, can be expressed by

$$B_{gm}=B_{st}\frac{P_g}{\mu_{st}}\,\frac{\mu_p+\mu_{st}}{P_g+P_i+\mu_p P_m} \qquad (1)$$

where the slope μ_{st} of the stabilizing shearing line is

$$\mu_{st}=\frac{P_i+P_g}{P_m}\,\frac{P_i+kz_{sc}}{P_i+P_g+kz_{sc}} \qquad (2)$$

and the constant k is

$$k=\frac{945C_p m}{C_m C_1 k_d^2(n_e/100)^2(\text{rpm})/1{,}000} \qquad (3)$$

Three-phase short-circuit stabilization of 2-pole machines frequently results in

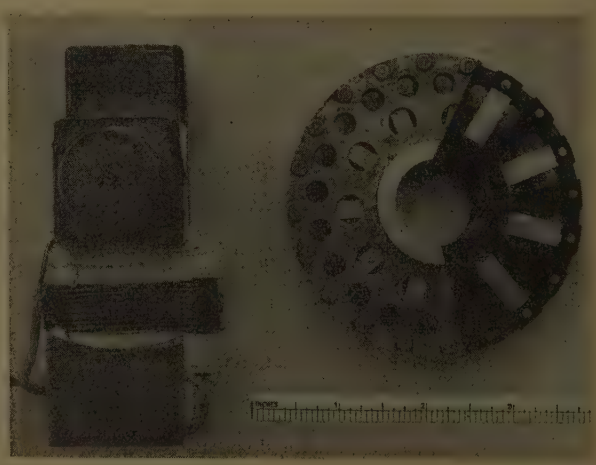


Fig. 4. Twenty-pole rotor with magnetizer which produces unidirectional radial magnetizing field. Alternator is 100 watts, 4,000 cycles per second, 24,000 rpm, designed to be mounted on a through shaft; rotor weight = 12 ounces

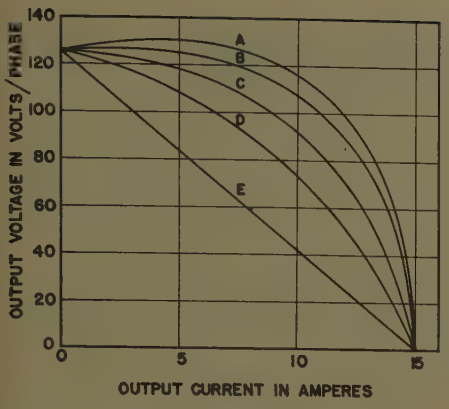


Fig. 5. Voltage-versus-current characteristic curves for a 1.5-kva 3-phase permanent-magnet alternator. Power factors:

A—0.95 leading
B—Unity
C—0.95 lagging
D—0.80 lagging
E—Zero lagging

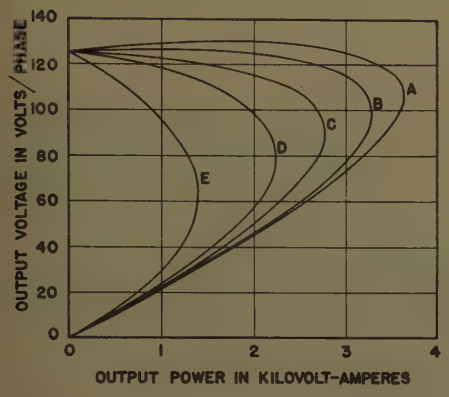


Fig. 6. Voltage-versus-power characteristic curves for a 1.5-kva 3-phase permanent-magnet alternator. Power factors:

A—0.95 leading
B—Unity
C—0.95 lagging
D—0.80 lagging
E—Zero lagging

a low stabilization point, because of the low in-stator permeance P_t. In such cases, it is recommended that the poles be provided with an additional leakage field, as shown in Fig. 2. The magnetic stabilization by short circuit renders the output voltage independent of demagnetizing effects of switching operations. This advantage, however, is only obtainable at the expense of increased weight.

For example, if the rotor has been magnetized in the stator, the initial working point would be located at the intersection of the line $\mu_{st}=(P_t+P_g)/P_m$, with the demagnetization curve. Application of load would bring the origin of the minor hysteresis loop to a lower point on the demagnetization curve. The exact location of this point depends upon the magnitude and the phase angle of the current. In the case of stabilization by short circuit, it is obvious from equation 2 that the stabilizing point on the demagnetizing

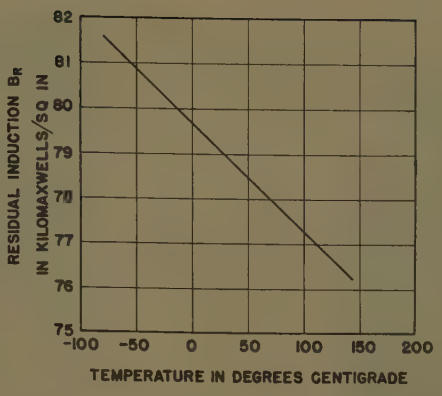


Fig. 7. Variation with temperature in residual induction of typical Alnico V permanent magnet

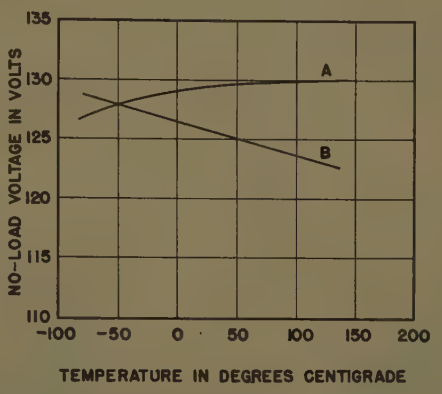


Fig. 8. Magnetic temperature compensation. No-load voltage over a temperature range of:

A—Overcompensated rotor
B—Noncompensated rotor

curve moves downward until the aforementioned condition of balances between magnetomotive forces and between voltages is reached. Thus, the useful air-gap flux and, consequently, the voltage output will be reduced. To maintain the rated output characteristic of a machine not stabilized, an increase of the rotor volume will be necessary.

If a sudden short circuit is applied, the subtransient component of the short-circuit current will decrease the stabilizing point below the value expressed by equation 2. The degree of the change depends upon the leakage reactance of the damper winding, which should be kept at a minimum. A carefully designed aluminum structure, completely surrounding the magnets, can reduce the effect of the transient demagnetization to less than 3% of the output voltage of a machine stabilized by steady-state short-circuit current.

Air Stabilization

Air stabilization is used in applications where frequent disassembly of the alternator is required without special tools. An

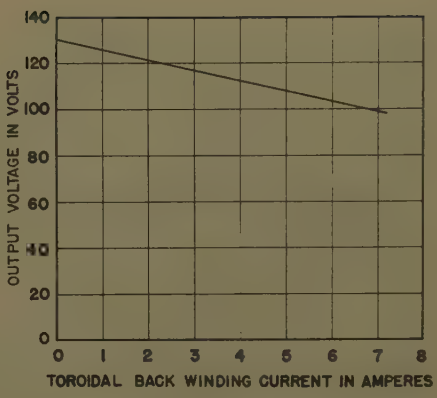


Fig. 9. Characteristic curve for typical toroidal back winding of no-load output voltage versus control direct current

air-stabilized rotor can be readily removed from the stator without a special keeper and yet maintain its output voltage after reassembly. A size and weight increase of 20 to 40% can be expected for an air-stabilized alternator having the same output characteristics as one which has been short-circuit stabilized.

Voltage Regulation and Control

Voltage control of small high-speed permanent-magnet alternators is a design problem because, in general, conventional types of control equipment have not been decreased in size and weight as rapidly as the alternators. Where feasible, the first design objective is to utilize the inherent regulation of the alternator. This approach eliminates all external controls. However, for lagging power-factor loads, or where close voltage control over a broad output range is required, the weight of the alternator with inherent regulation may become excessive, making external controls necessary. Figs. 5 and 6 show typical inherent regulation curves, based on a nominal rating which varies in accordance with the demands of a given application. Close inherent voltage con-

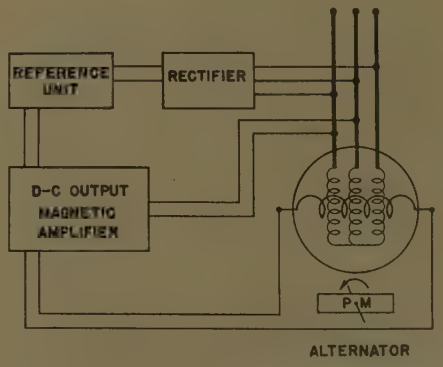


Fig. 10. Automatic control system using toroidal back winding for voltage control of permanent-magnet alternator

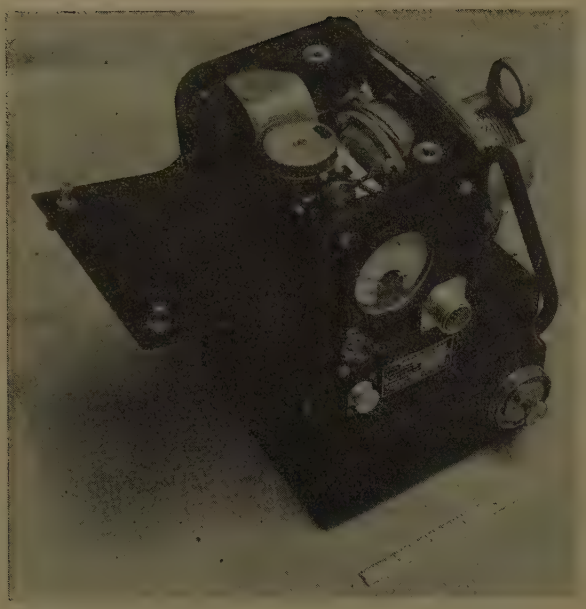


Fig. 11 (left). Pockette power plant; a miniature 100-watt 800-cycle-per-second 12,000-rpm semidiesel engine-generator unit. Weight 12 pounds including fuel for 8 hours

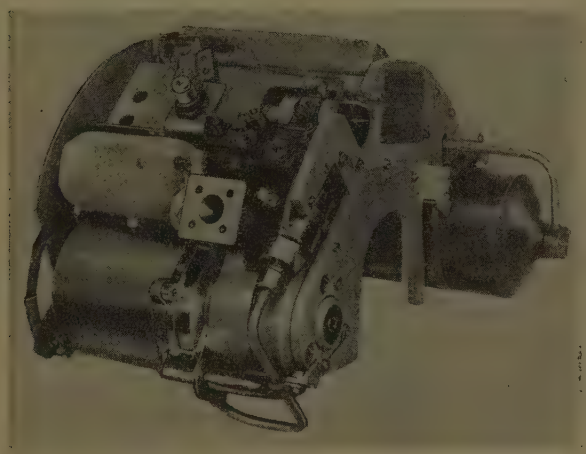


Fig. 12 (below, left). Accessory power supply; 4.0-horsepower 650-watt 400-cycle-per-second 24,000-rpm liquid-monopropellant-fuel hot-gas-turbine-alternator unit. Weight 47 pounds including fuel for a 60-second run

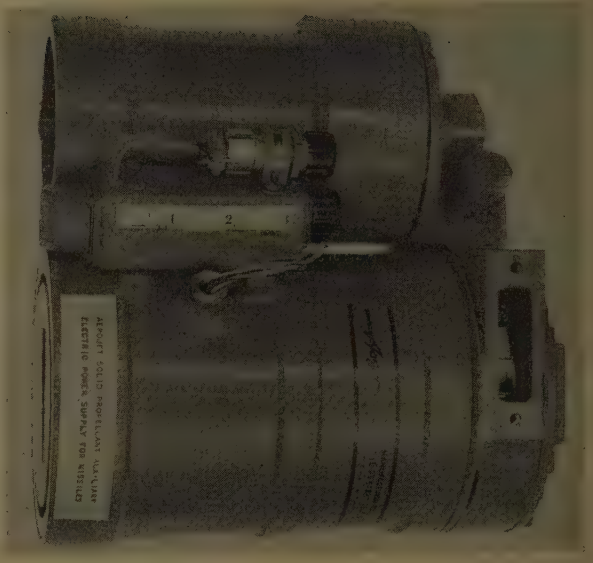


Fig. 13. Auxiliary power unit; 500-watt 400-cycle-per-second 12,000-rpm solid-propellant hot-gas-turbine-alternator unit. Weight 40 pounds with fuel for a 90-second run

Fig. 14 (right). Auxiliary power unit; 1,500-watt 400-cycle-per-second 24,000-rpm compressed-gas-turbine-alternator unit. Weight 17 pounds (dry)

trol can be maintained over a maximum load range by: 1. establishing unity, or slightly leading, power-factor loads, or 2. increasing the power factor as step loads are applied. In the salient-pole alternator, temperature compensation and toroidal back winding are used for voltage control.

Temperature Compensation

The output voltage of a permanent-magnet alternator using *Alnico V* magnets increases with decreasing temperature. This increase is of the order of 6% over the temperature range from +150 degrees centigrade to −73 degrees centigrade, corresponding to an equivalent increase in the residual induction (B_r) of the magnet (Fig. 7). In applications where the inherent voltage regulation must be maintained within close limits, the alternator is temperature-compensated to eliminate this voltage change.

The temperature compensation is accomplished by substituting a temperature-sensitive (magnetic) metal for a portion of the stainless-steel (nonmagnetic) plates used in the rotor-cage structure. The function of this temperature sensitive ("thermomagnetic") alloy is to control, or shunt, a given amount of the usable generating flux. By providing a magnetic shunt made of a carefully selected thermomagnetic alloy[5], it is possible to compensate fully for the foregoing temperature effects. An alternator so compensated delivers a voltage which is essentially independent of temperature.

In practice, it is actually preferable to overcompensate in order to allow for the decrease in shunt resistance of the stator winding at low temperatures. Fig. 8 compares the no-load output voltage of an overcompensated rotor with that of a noncompensated rotor over a given temperature range.

Toroidal Back Winding

In applications where the inherent voltage of the alternator cannot be held within the required limits, a toroidal back winding is used for voltage control. This control winding is wound as a uniform-series toroidal winding around the armature core. The application of direct current to this winding causes a unidirectional field to be superposed on the alternating field in the armature core. As the superposed direct field is increased, the reluctance of the core increases, thus lowering the generated voltage. As shown in Fig. 9, the change in generated voltage is practically linear with the applied control current. Maximum control excitation is required under minimum or no-load conditions. Consequently, there is no increase of the armature heating, or decrease in the efficiency, under full load, as is the case in conventional alternators with field windings.

A typical automatic control system using the toroidal back winding is shown in Fig. 10. Static control systems of this type have been used to control the average

value of the 3-phase output voltage within 1% over the complete load range.

Design of Toroidal Back Winding

It has been shown in the case of an air-stabilized machine,[2] that the density of the portion of the permanent-magnet flux crossing the air gap can be expressed by the equation

$$B_{gm}=B_{st}\frac{P_g}{P_o}\frac{P_o+\mu_p P_m}{P_g+P_t+\mu_p P_m} \quad (4)$$

where the air-gap permeance is

$$P_g=3.19\frac{\pi l}{2}\frac{C_p d_\tau}{g_e p} \quad (5)$$

The value g_e in equation 5 is a function of the air gap and the equivalent path of all soft-iron components of the magnetic circuit. Generally, the latter is so designed as to work essentially in the linear range of its magnetic characteristic.

If a direct magnetic field is superposed on the alternating field in the core of the stator, the permeance of the iron path will decrease. The ratio of the permeance of the air gap to the permeance of the stator core may be expressed as

$$\tau=P_g/P_b \quad (6)$$

where the permeance P_b of the stator core can be calculated by the equation

$$P_b=\frac{6.38}{\pi}\mu_e\frac{h\sigma l p}{D-h} \quad (7)$$

where μ_e represents the effective permeability of the stator core iron.

Substituting equations 5 and 7 in equation 6 yields

$$\tau=2.45\frac{C_p d_\tau(D-h)}{\mu_e g_e h\sigma p^2} \quad (8)$$

Replacing the air-gap permeance P_g in equation 4, by the equivalent permeance of air gap and stator core, $P_{ge}=P_g/(1+\tau)$, results in the reduced air-gap flux density at no load of

$$B_{gm}'=B_{st}\frac{P_g}{P_o}\frac{P_o+\mu_p P_m}{P_g+(1+\tau)(P_t+\mu_p P_m)} \quad (9)$$

Similar, for the case of short-circuit stabilization

$$B_{gm}=B_{st}\frac{P_g}{\mu_{st}}\frac{\mu_p+\mu_{st}}{P_g+(1+\tau)(P_t+\mu_p P_m)} \quad (9A)$$

The introduction of the effective permeability μ_e in equation 7 necessitates an explanation.

Upon entering the core, the rotor field subdivides in two opposite directions. Since the superposed direct field is unidirectional, the total field must be either the sum or the difference of the two, as the case may be. However, due to the nonlinear magnetic characteristic of the soft iron core, one half-wave of the resultant (polarized and synchronously rotating) flux is peaked and the other half-wave is flat-topped.[6] In addition, the rotating field in the air gap is far from sinusoidal in shape. Consequently, the required values of effective permeability must be determined from tests on a typical machine. The values thus obtained can be applied to design calculations of other machines if their core material is identical and the general geometrical configuration of their magnetic circuit is similar to that of the test machine.

The vectorial sum of the induced voltages in the conductors of the back winding is zero under ideal conditions. Irregularities in the winding and small unsymmetries in the magnetic circuit, however, may cause the build-up of a residual alternating voltage on the terminals of the direct back winding. This fact must be considered in connection with the design of the control circuit.

Applications

As a component in various types of power supplies, the salient-pole permanent-magnet alternator provides simple, rugged construction and the following significant advantages: 1. reliability with minimum weight, 2. high efficiency, 3. positive excitation, and 4. the absence of external excitation, which eliminates problems associated with moving contacts. Typical fields where these characteristics are required are in guided-missile auxiliary-power units, ground-support power supplies, emergency power supplies, and electric conversion units.

Fig. 11 shows a miniature, semidiesel engine-generator unit, using a small permanent-magnet alternator. When coupled directly to the engine crankshaft as a flywheel, a rugged and compact alternator rotor construction is essential.

A liquid-monopropellant-fuel hot-gas-turbine-alternator power supply is shown in Fig. 12. With the rotor directly coupled to the turbine, miniaturization is realized through high-speed operation. Operating on a monopropellant fuel, this unit provides electric and mechanical power completely independent of the missile's main power plant and fuel system. By temperature compensating the alternator rotor, the electric output voltage is held relatively constant over the extreme temperature conditions imposed by high-speed and high-altitude flight.

Another hot-gas turbine-alternator power unit, which uses a solid propellant, is shown in Fig. 13. This unit is a lightweight source of electric energy for airborne applications. It consists essentially of a solid rocket propellant, an impulse turbine, an alternator, and an electromagnetic control system able to withstand long-term storage and insensitive to vibration, shock, and all normal air-borne environmental conditions.

A compressed-gas turbine-alternator

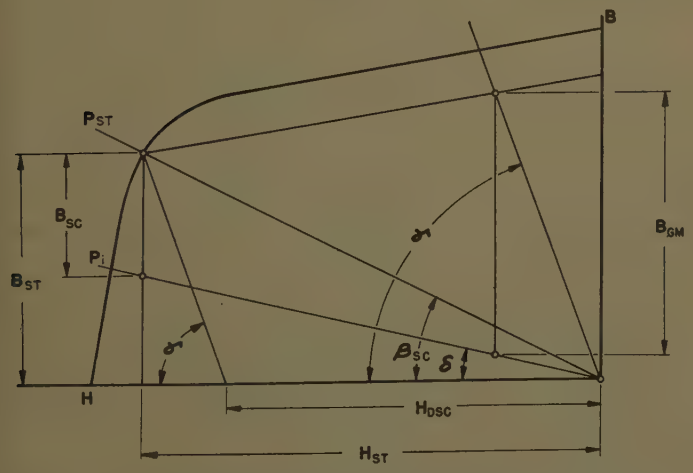


Fig. 15. Graphic representation of permanent-magnet circuit for stabilization by short circuit

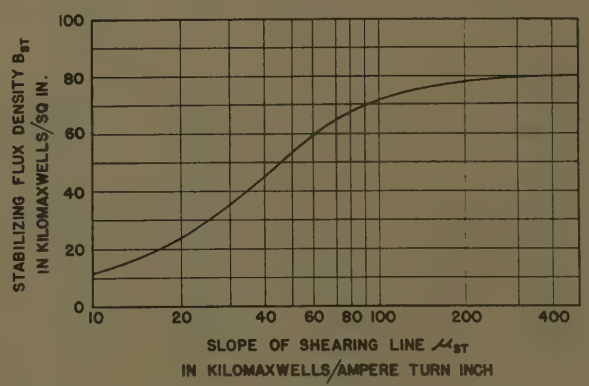


Fig. 16. Demagnetizing curve for Alnico V, represented as function of magnet flux density and slope of shearing line

power supply is shown in Fig. 14. Operational features of this power supply consist of close regulation of the speed and output voltage through the use of an electromagnetic speed-control system and the automatic toroidal back winding voltage-system.

Appendix

Air-Gap Flux Density for Short-Circuit Stabilized Machine

Fig. 15 represents the geometry of the permanent-magnet circuit.[2] It is assumed that the permanent magnets are stabilized on the demagnetization curve at the point B_{st}, H_{st}. The distance from this point to the intersection of its ordinate with the leakage line P_t is proportional to the air-gap flux density at short circuit B_{sc}.

Then, from Fig. 15

$$B_{sc}=B_{st}-H_{st}\tan\delta \tag{10}$$

and the magnet flux density at the stabilization point is

$$B_{st}=(H_{st}-H_{dsc})\tan\gamma \tag{11}$$

where H_{dsc} represents the demagnetizing stator ampere-turns per-inch magnet length in short circuit.

Substituting

$$\mu_{st}=\frac{B_{st}}{H_{st}} \tag{12}$$

in equations 10 and 11 gives

$$B_{sc}=B_{st}(1-\tan\delta/\mu_{st}) \tag{13}$$

and

$$B_{st}=H_{dsc}\mu_{st}\tan\gamma/(\tan\gamma-\mu_{st}) \tag{14}$$

The hypothetical total magnetic flux in the air gap can be expressed by the equation[5]

$$\phi_r=B_g d_r\pi l=\frac{6{,}000E_t10^6}{C_w n_e(\text{rpm})} \tag{15}$$

where

$$C_w=\frac{E_t C_1 k_d}{E_p\sqrt{2}m} \tag{16}$$

In the case of symmetrical short circuit ($E_p=E_{sc}$), the flux density B_{sc} corresponds to the remaining air-gap flux density, reduced to the magnet area,

$$B_{sc}=B_g\frac{C_p d_r\pi l}{b_p l_p p} \tag{17}$$

Thus, from equations 15 through 17

$$B_{sc}=\frac{8{,}450C_p m10^6}{b_p l_p p C_1 k_d n_e(\text{rpm})}E_{sc}=k_1E_{sc} \tag{18}$$

The induced emf E_{sc} must balance the impedance drop at short circuit

$$E_{sc}=\sqrt{r^2+X_{sc}^2}\,I_{sc}=z_{sc}I_{sc} \tag{19}$$

The armature reaction can be expressed by[1]

$$F_{dm}=0.45n_e I_{sc}C_m k_d/p \tag{20}$$

Substituting $F_{dm}=H_{dsc}h_p$, the stator current can be written

$$I_{sc}=h_p p/(0.45n_e C_m k_d)H_{dsc}=k_2H_{dsc} \tag{21}$$

Combining equations 18, 19, and 21 yields

$$B_{sc}=k_1k_2z_{sc}H_{dsc} \tag{22}$$

Equating equations 13 and 22 and substituting equation 14 gives the slope of the stabilizing shearing line at short circuit

$$\mu_{st}=\frac{\tan\delta\tan\gamma+k_1k_2z_{sc}\tan\gamma}{\tan\gamma+k_1k_2z_{sc}} \tag{23}$$

Introducing the following equation

$$k=k_1k_2P_m=\frac{945C_p m}{C_m C_1 k_d^2(n_e/100)^2(\text{rpm})/1{,}000} \tag{24}$$

and substituting

$$\tan\delta=P_t/P_m \tag{25}$$

$$\tan\gamma=P_t/P_m+P_g/P_m \tag{26}$$

gives

$$\mu_{st}=\frac{P_t+P_g}{P_m}\,\frac{P_s+kz_{sc}}{P_t+P_g+kz_{sc}} \tag{27}$$

The stabilizing flux density B_{st} can be found directly for *Alnico V* from Fig. 16 after computing μ_{st} from equation 27.

Equation 4 has been derived on the basis of air stabilization at the outstator permeance, $P_o=P_m\tan\beta$. For stabilization at short circuit, equate $P_o=\mu_{st}P_m$ and obtain

$$B_{gm}=B_{st}\frac{P_g}{\mu_{st}}\,\frac{\mu_p+\mu_{st}}{P_g+P_t+\mu_p P_m} \tag{28}$$

References

1. Design Calculations for A-C Generators, **David Ginsberg.** *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 1274–80.

2. Synchronous Machines with Rotating Permanent-Magnet Fields, Part II. Magnetic and Electrical Design Considerations, **Fritz Strauss.** *Ibid.*, vol. 71, pt. III, Oct. 1952, pp. 887–93.

3. Synchronous Machines with Rotating Permanent-Magnet Fields, Part I. Characteristics and Mechanical Construction, **Maurice W. Brainard.** *Ibid.*, Aug. 1952, pp. 670–76.

4. Rotor Construction and Method of Magnetizing the Same, **M. W. Brainard.** *U. S. Patent no. 2,754,440*, July 1956.

5. Temperature-Compensator Alloys, **Warren S. Eberly.** *Machine Design*, Cleveland, Ohio, vol. 26, May 1954, pp. 152–56.

6. Alternating Current Wave Forms (book), **Philip Kemp.** Chapman & Hall, Ltd., London, England, 1952, pp. 279–81.

A Method of Fault Analysis for A-C Aircraft Electric Power Systems

L. J. RINDT
NONMEMBER AIEE

R. D. JESSEE
ASSOCIATE MEMBER AIEE

THE increasing electric power requirements in today's aircraft make necessary the continual search for better methods of protection against faults which may cause the loss of electric power or otherwise endanger the mission. Power-system studies have been made in the past using a mock-up of the electric system composed of actual components, and have proved quite successful. However, it is desirable to obtain information faster, and before the development of complete apparatus. Parameters such as line voltages and currents, positive-, negative-, and zero-sequence voltages and currents must be known in order effectively to apply system protection, and they are of much value in regulator design. Faults occurring where they produce the same effect on all regulators, such as at the bus, can be calculated very simply by the method of symmetrical components. A means for detecting such faults by use of negative-sequence voltage is pointed out in another paper.[1] However, in multi-generator systems, load-division circuits are employed in order that the load current may be properly divided between generators. Generator currents are usually sensed in a single phase through a current transformer; consequently, generators are forced to divide load current relative to the current in the current transformer associated with a particular generator feeder. Should a fault occur on the load-division phase at a point between a generator and its load-division current transformer, as shown in Fig. 1, that generator would become overexcited and supply practically all of the fault current.

Paper **57-479**, recommended by the AIEE Air Transportation Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE East Central and Middle Eastern District Meeting and Air Transportation Conference, Dayton, Ohio, May 7–9, 1957. Manuscript submitted February 1, 1957; made available for printing May 17, 1957.

L. J. Rindt and R. D. Jessee are with the Westinghouse Electric Corporation, the former at East Pittsburgh, Pa., and the latter at Lima, Ohio.

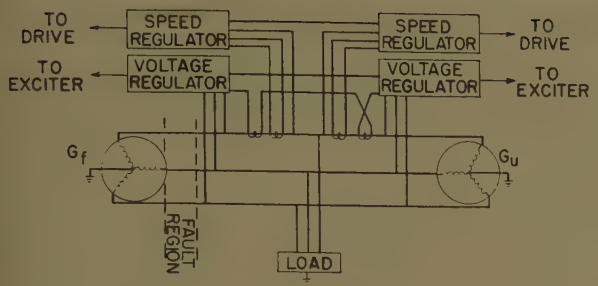


Fig. 1. Diagram of regulated system

Such a condition would be very difficult to calculate by ordinary methods.

The method of analyzing this type of system as presented here is to set up passive elements of an a-c calculating board to represent the elements of the power system and to use an electronic differential analyzer to represent the regulators of this system. Although there are transients which occur between the time the regulators are told to change until they have completed their change, the final steady-state conditions are the significant ones. At this steady-state point, the magnitudes and phase angles of the positive-, negative-, and zero-sequence components of the generator currents and voltages are needed. From these values, the generator power, voltages, and line currents can be computed.

In addition to evaluating this method, the effects of faults on a power system may be analyzed. With various system loadings, the effects of the following faults have been analyzed.

1. Single line-to-ground faults: The fault location is on the load-division phase at a point between the generator terminal and the load-division current transformer, as indicated in Fig. 1. The regulator receives its voltage signal from the load bus. Several different values of fault impedances are analyzed but in all cases the fault impedance is purely resistive.

2. Line-to-line faults: The fault location and impedances are the same as in the case of the single-line-to-ground fault. One of the faulted lines is the load-division phase.

3. Open-phase faults: The fault location is at a point between a generator and the load bus. This type of fault is applied to each phase with various amounts of load.

An analog-computer study has been undertaken to obtain data necessary to analyze these and evaluate the method.

Assumptions and Limitations

To set up the power system problem for analysis it is necessary to make certain assumptions and define limits imposed on certain parameters.

1. Saturation is neglected.

2. The exciter is assumed to be capable of supplying generator excitation demanded by the regulator sensing circuit from zero to ceiling.

3. Ceiling excitation voltage is assumed to be the same for all conditions.

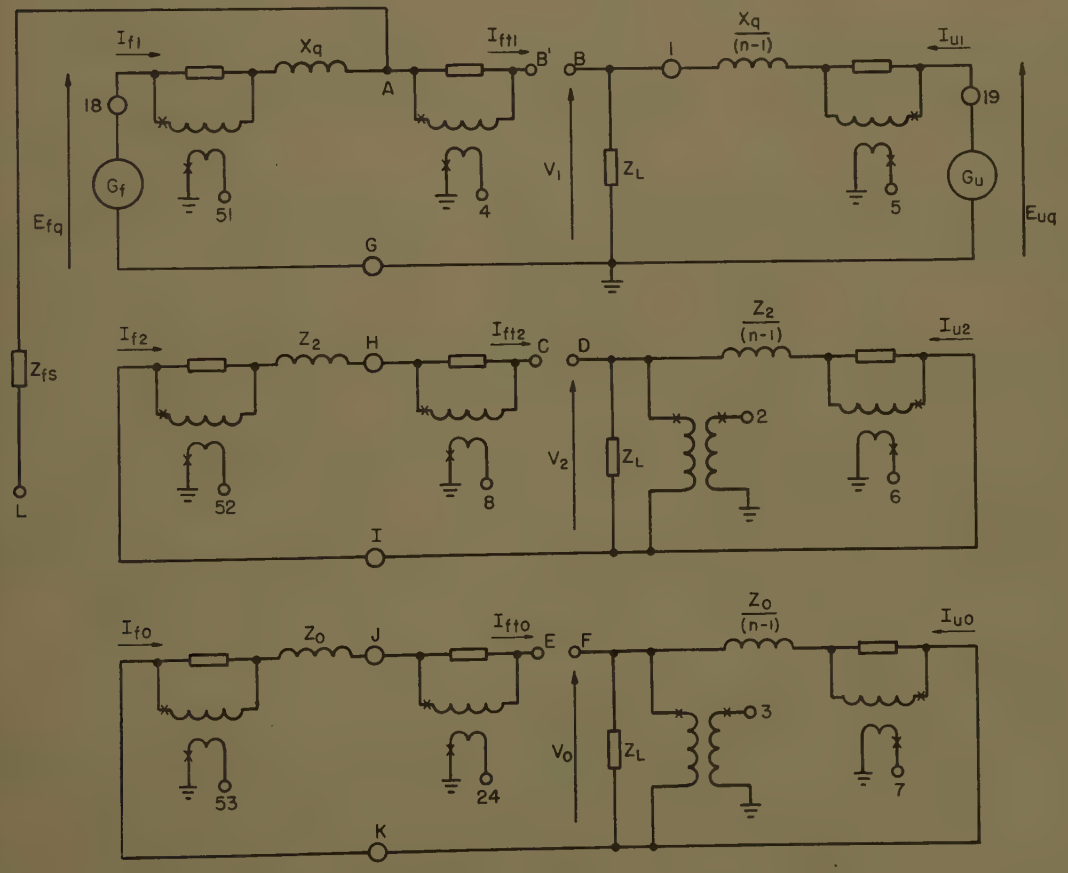
4. The generator is driven through an overrunning clutch which does not allow the prime mover to take power from the electric power system. It is, therefore, assumed that the maximum power taken from the system is 0.05 per-unit power, required to drive the clutch.

5. In any case where a generator is being motored and sufficient power cannot be supplied from the line to maintain synchronism, the generator is replaced by a constant impedance.

6. Frequency and temperature are assumed to be constant.

Nomenclature

G_u = unfaulted generator; represents $(n-1)$ generators
G_f = faulted generator
n = number of generators in system
E_{fq} = quadrature voltage of G_u
E_{uq} = quadrature voltage of G_f



TYPE OF FAULT	SHORT
SINGLE LINE TO GROUND	B' TO B C TO D E TO F G TO H I TO J K TO L
LINE TO LINE	B' TO B C TO D E TO F G TO I H TO L
OPEN PHASE	B' TO C TO E B TO D TO F

Fig. 2. Three sequence networks of power system

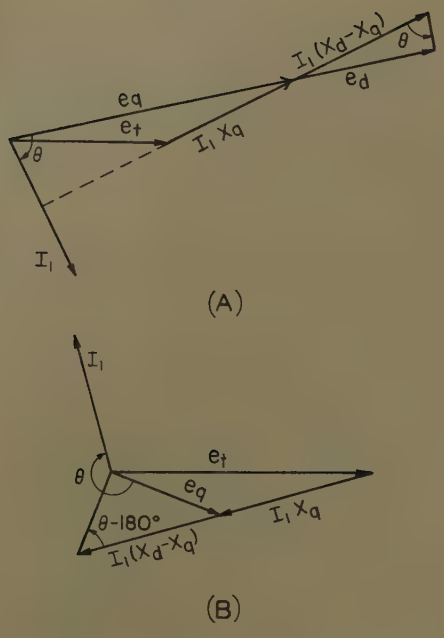


Fig. 3. Excitation phasor diagram

A—Maximum excitation
B—Minimum excitation

e_q = quadrature voltage of generator
e_d = excitation voltage of generator
e_t = terminal voltage of generator
V_1 = positive-sequence voltage at load bus
V_2 = negative-sequence voltage at load bus
V_0 = zero-sequence voltage at load bus
V_A, V_B, V_C = load-bus phase voltages
E_{f0} = reactive load-division correction voltage for G_f
E_{u0} = reactive load-division correction voltage for G_u
E_{fr} = regulated reference voltage for G_f
E_{ur} = regulated reference voltage for G_u
I_{f1} = positive-sequence current of G_f
I_{f2} = negative-sequence current of G_f
I_{f0} = zero-sequence current of G_f
I_{ft1} = positive-sequence current of G_f current transformer
I_{ft2} = negative-sequence current of G_f current transformer
I_{ft0} = zero-sequence current of G_f current transformer
I_{u1} = positive-sequence current of G_u
I_{u2} = negative-sequence current of G_u
I_{n0} = zero-sequence current of G_u
I_{fa}, I_{fb}, I_{fc} = phase currents of G_f
I_{fta} = phase A current of G_f current transformer
I_{ua}, I_{ub}, I_{uc} = phase currents of G_u
I_1 = positive-sequence current of generator
Z_q = quadrature reactance of generator
Z_2 = negative-sequence impedance of generator
Z_0 = zero-sequence impedance of generator
Z_L = load impedance
Z_{fs} = sequence representation of fault impedance
X_d = direct-axis reactance of generator
X_q = quadrature-axis reactance of generator
θ = angle between e_q and I_1
M = voltage-regulator constant of per-unit voltage/per-unit current

All sequence current and voltages refer to phase A quantities.

Representation of Power System

In this method of analysis, the power-system portion is set up on a single-phase basis using the three sequence networks, as shown in Fig. 2. For simplicity and because of their smallness, the line impedances are neglected in the case presented here. However, they may be added if their size indicates they will have a significant effect on the results. The generators G_f and G_u supply the voltages E_{fq} and E_{uq}, behind quadrature reactance. This particular representation was chosen because of two factors: 1. The generator can be represented by a single reactance and a single voltage source. 2. The voltage source used has the same phase angle as the excitation voltage, thus simplifying the calculations needed to determine the excitation voltage of the salient-pole machine.[2] A phasor diagram showing the relationship of the different voltage and voltage drops is shown in Fig. 3, and the method used to evaluate the excitation voltage is discussed under "Results."

As is often done in power-system analysis, the machine saturation is neglected. This will have no effect on the results where the system regulates within the range of the excitation voltage, but it will have some effect on the results where the upper limit of this range is reached. However, it is felt that this effect is not important enough to justify the additional refinement needed to take saturation into account.

The various connections that have to be made to represent the different faults are shown in Fig. 2. The currents and voltages which need to be measured and fed to the electronic differential analyzer are also shown here, and are read with respect to ground. To make this possible transformers are inserted in the negative- and zero-sequence networks to isolate the load-bus voltage from ground. Since the electronic differential analyzer operates from voltages, suitable dropping resistors and isolating transformers are inserted to give voltage drops proportional to the currents to be measured. As can be seen from the polarity markings on the transformers, points *5*, *6*, and *7* represent positive currents, while points *4*, *8*, *24*, *51*, *52*, and *53* represent negative currents. The reason for this is made apparent in the appendix.

Using a 40-kva base in this study, the basic per-unit values are as follows:

System Values	Board Values	Per-Unit Values
120 volts	10 volts	1.0
1.09 ohms	1,000 ohms	1.0
111 amperes	0.01 ampere	1.0

Representation of Regulators

The regulators represented here are both a speed and a voltage regulator. The speed regulator makes use of a single-phase voltage and a single-phase current for real load division. The voltage regulator uses the 3-phase voltages and the single-phase current from each generator through a reactive-load equalizing circuit. A description of these regulators appears in the appendix.

Procedure

In general, the procedure used in each of the studies is as follows: With no load or fault on the system, both generator voltages are brought in phase and set equal in magnitude to 1.0 per unit. This condition gives a bus voltage of 1.0 per unit with no current flowing between generators. Under this condition, voltage references E_{fr} and E_{ur}, Fig. 4(E), which the voltage regulators attempt to hold constant, are established. As a check, the speed regulator should indicate zero. These reference points are used througout all of the studies to indicate whether a regulated condition exists. The next step is to add the required load to the system. The voltages of the two generators are then readjusted until the two voltage-regulator signals have returned to their original reference points. Since the load is applied to the bus, and the generator voltages are initially in phase, no phase-angle adjustment is required.

The next step is to apply the fault to the system and to incrementally adjust the voltages and phase angles of the generators until the regulator signals return to their original reference. This return to the original reference in some cases will not be complete because of other limitations of the system which occur first. These limitations are the maximum and minimum excitation voltages, and a maximum amount of power that can be taken into any machine. These limitations are discussed more in detail in "Results."

When the original references are reached and/or the other limitations satisfied, a steady-state condition exists in the system. At this point readings are taken and the pertinent points recorded.

Results

Although several types of faults have been analyzed, the results discussed here are confined to the case of a single line-to-ground fault, which is presented as an example. As many deduced from the system diagram, Fig. 1, faults occurring on

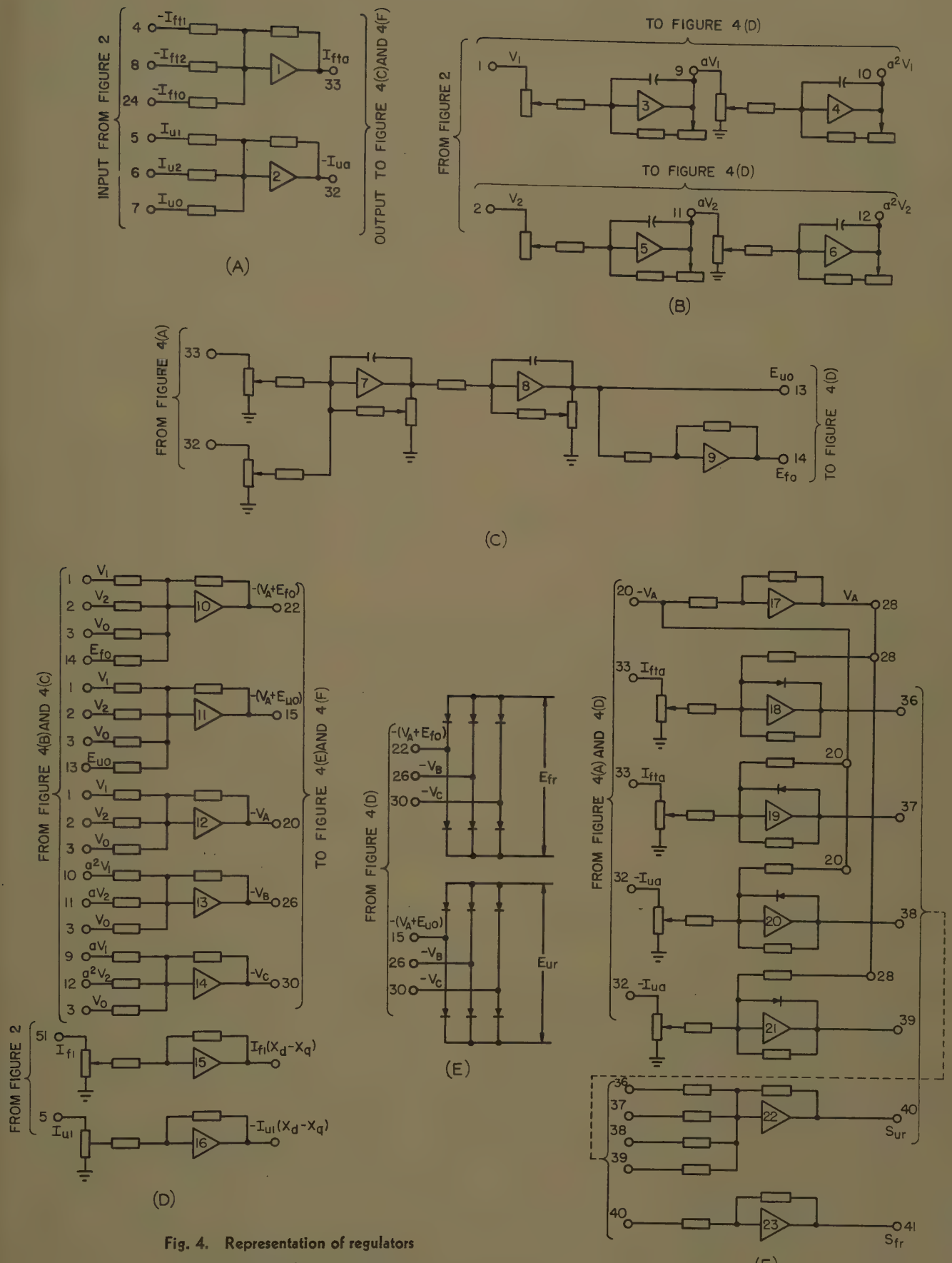


Fig. 4. Representation of regulators

A—Phase currents from sequence currents
B—120- and 240-degree rotation of sequence voltages
C—Voltage-regulator correction signals
D—Corrected phase A voltage, phase voltages from sequence voltages, excitation component
E—Voltage-regulator signal to exciter
F—Speed-regulator signal

Table I. Line-to-Neutral Fault on 2-Generator System With No Load*

Value	Z_{fs} = †3.0	Z_{fs}† = 1.0	Z_{fs}† = 0.6	Z_{fs}† = 0
E_{fq}	1.03 ∠ 42.3	1.23 ∠ 58.2	1.43 ∠ 67.6	2.1 ∠ 27
E_{uq}	0.985 ∠ 27	0.855 ∠ 27	0.705 ∠ 27	
V_1	0.995 ∠ 25.1	0.935 ∠ 25.8	0.855 ∠ 27.2	0.225 ∠ 1.0
V_2	0.0252∠273.7	0.068 ∠270.5	0.100 ∠268.5	0.190 ∠187.3
V_0	0.0035∠261.8	0.0096∠260	0.0142∠258	0.027 ∠176
V_A	0.99 ∠ 25	0.91 ∠ 22.2	0.815 ∠ 20.6	0.0225∠292.3
V_B	0.98 ∠267	0.905 ∠270	0.92 ∠273.5	0.345 ∠265
V_C	1.01 ∠147	0.995 ∠148.4	0.95 ∠149.6	0.38 ∠100
I_{f1}	0.06 ∠354.5	0.136 ∠318.8	0.236 ∠312.5	0.396 ∠316.7
I_{f2}	0.164 ∠207	0.456 ∠202.7	0.66 ∠200.2	1.27 ∠115.2
I_{f0}	0.168 ∠206.5	0.460 ∠203	0.66 ∠200.6	1.27 ∠115.8
I_{u1}	0.06 ∠174.5	0.136 ∠138.8	0.236 ∠132.5	0.396 ∠136.7
I_{u2}	0.164 ∠ 27	0.456 ∠ 22.7	0.66 ∠ 20.2	1.27 ∠295.2
I_{u0}	0.168 ∠ 26.5	0.460 ∠ 23	0.66 ∠ 20.6	1.27 ∠295.8
I_{f1}	0.3o4 ∠ 24	0.980 ∠ 16.5	1.46 ∠ 12.8	2.86 ∠299.3
I_{f2}	0.176 ∠ 26	0.466 ∠ 22.7	0.68 ∠ 20.7	1.3 ∠296.5
I_{f0}	0.172 ∠ 25.5	0.462 ∠ 22.7	0.67 ∠ 20.3	1.28 ∠296.2
	Regulates easily; unfaulted generator motors 2 kw	Regulates easily; unfaulted generator motors 2 kw	Regulates easily; unfaulted generator motors 2 kw	Unfaulted generator's excitation goes to zero and machine is replaced by an induction-motor equivalent of 0.4+j0.4 per-unit ohm. Faulted generator's excitation goes to maximum

* All values are per-unit quantities.

† For this case, Z_{fs} is equal to three times the fault impedance.

the load division phase cause generator G_f to supply practically all of the power absorbed in the fault. Table I shows the values of sequence voltages and currents obtained for the line-to-neutral fault on a 2-generator system at no load with various values of fault impedance. Table II shows similar results for a 2-generator system initially supplying 1/2 rated load. From the data contained in these tables, curves may be constructed to show voltage and current characteristics desired. For example, generator phase currents and bus phase voltages are shown in Fig. 5 as functions of fault current.

The last lines of Table I and Table II give some remarks concerning the particular condition. The following is a brief discussion of these remarks.

1. The term "regulates easily" denotes those cases where the regulator signals, with a minimum of oscillation, returned to their original references, or approached these references within the other limitations imposed on the system. In the cases where any of the limitations were reached, these were also noted.

2. During the course of the incremental adjustment of the two generators, the minimum and maximum excitation voltage on each machine was checked. This excitation voltage was $e_d = e_q + I_1(X_d - X_q)\sin\theta$, the minimum being zero and the maximum 4.85 per unit. When either of these limits was reached, the excitation voltage was held constant but the other variables were changed until the original reference or some other limitation was reached.

3. Another of the limitations reached was the condition where a generator motored power. It was determined from the actual machine and drive that the maximum power intake by a generator would be 0.05 per-unit power to drive the clutch and the shaft. In many cases the speed regulator signaled one of the generators to slow down considerably, causing an increase in the phase angle between generated voltages. In the process of increasing this phase angle, a point was reached where the generator was taking in 0.05 per-unit power and the speed-regulator signal still called for a further increase in phase angle. At this point, the power intake to the generator was held constant while the voltages were regulated.

In some of the cases studied, the regulator signals called for the excitation voltage to go to zero and for a considerable increase in phase angle between generated voltages. The terminal voltage in these cases was quite small. An analytical investigation of the terminal and excitation voltages of the circuit involved showed that the bus could not supply the generator with enough power to drive the shaft and clutch. This indicated that the generator pulled out of step and operated as an induction motor. Even with the generator replaced by an induction-motor equivalent, the necessary power could be be taken from the system. It was decided that under these conditions the effect of changes in the induction-motor equivalent would be small on the total system. Also the representation of less than 200% slip would become quite involved. As a result, the induction-motor equivalent was set at a fixed value of 0.4+j0.4 per-unit impedance for all cases of this type.

The significance of the conditions studied is that these particular faults do not show the fault as an additional load to the regulator. In the single line-to-ground and line-to-line faults, the portion of the fault current supplied by the un-

Table II. Line-to-Neutral Fault on 2-Generator System With 1/2-Rated Load*

Value	Z_{fs}† = 3.0	Z_{fs}† = 1.0	Z_{fs}† = 0.6	Z_{fs}† = 0
E_{fq}	1.38 ∠ 54	1.7 ∠ 67.4	1.96 ∠ 72.5	2.03 ∠ 27
E_{uq}	1.18 ∠ 27	1.08 ∠ 27	0.945 ∠ 27	
V_1	0.975 ∠ 23.5	0.96 ∠ 26.4	0.895 ∠ 27	0.197 ∠359
V_2	0.0231∠269.6	0.066 ∠268.5	0.0995∠266.7	0.163 ∠184.6
V_0	0.0035∠260	0.0098∠259.6	0.0148∠257.7	0.0246∠175.3
V_A	0.98 ∠ 23	0.935 ∠ 22.5	0.85 ∠ 20.3	0.02 ∠291.4
V_B	0.97 ∠265	0.935 ∠269.5	0.865 ∠272.4	0.30 ∠262.1
V_C	1.0 ∠145	1.01 ∠148	0.985 ∠148.8	0.325 ∠ 97.6
I_{f1}	0.74 ∠359	0.84 ∠357	0.85 ∠350.2	0.536 ∠313.6
I_{f2}	0.174 ∠207	0.498 ∠205	0.74 ∠202.7	1.24 ∠117.3
I_{f0}	0.164 ∠205.3	0.474 ∠203.6	0.70 ∠201	1.17 ∠115.8
I_{u1}	0.316 ∠296.4	0.21 ∠281.8	0.095 ∠260	0.344 ∠133.5
I_{u2}	0.156 ∠ 22.5	0.454 ∠ 20.8	0.67 ∠ 18.5	1.15 ∠293
I_{u0}	0.163 ∠ 23.5	0.466 ∠ 22.5	0.69 ∠ 20.5	1.16 ∠295.1
I_{f1}	1.08 ∠ 7.3	1.75 ∠ 11.4	2.16 ∠ 10	2.8 ∠302
I_{f2}	0.162 ∠ 21.7	0.458 ∠ 20.5	0.67 ∠ 18.3	1.13 ∠293.5
I_{f0}	0.167 ∠ 23.1	0.474 ∠ 22.1	0.69 ∠ 20	1.16 ∠295.3
	Regulates easily	Regulates easily; unfaulted generator motors 2 kw	Regulates easily; unfaulted generator motors 2 kw	Unfaulted generator's excitation goes to zero and machine is replaced by induction-motor equivalent of 0.4+j0.4 per-unit ohm. Faulted generator's excitation goes to maximum

* All values are per-unit quantities.

† For this case, Z_{fs} is equal to three times the fault impedance.

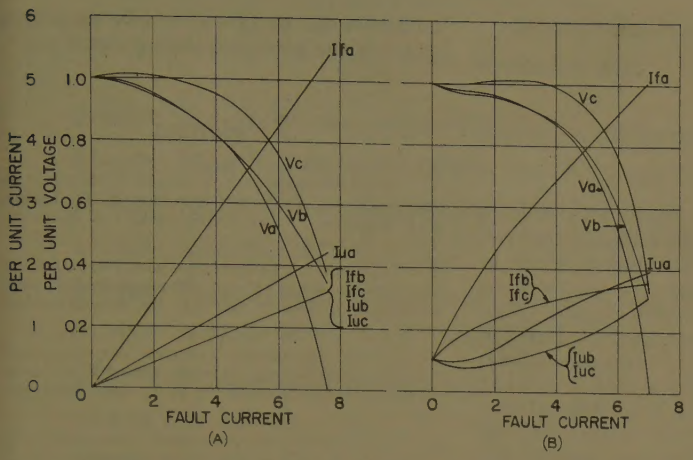


Fig. 5. Generator current and voltage characteristics with line-to-neutral fault on 2-generator systems

A—No load B—1/2-rated load

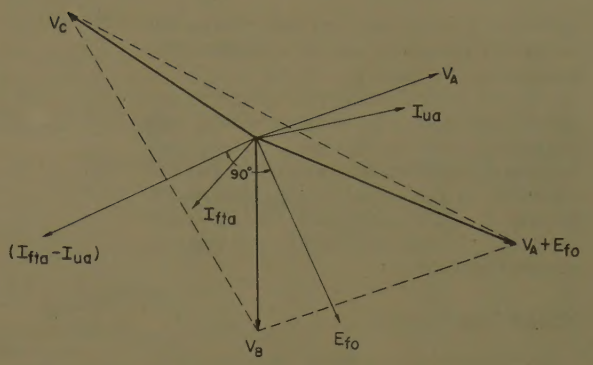


Fig. 6. Voltage-regulator phasor diagram

faulted machine G_u appears as current flowing into the faulted machine G_f. This effect, in general, causes the faulted machine to tend to speed up and supply nearly all of the fault current. As load was added to the bus, there was a tendency for the load current to be shared by the generators. In this case, and with a high-fault impedance, the unfaulted generator carried some load. With lower values of fault impedance, the unfaulted generator tended to slow down until it was motoring.

Fault current flowing through the current transformer of the faulted machine in a reversed direction indicated to the faulted machine that it was taking in reactive power. This caused the faulted machine to increase its excitation considerably. In many cases the faulted machine reached maximum excitation.

Conclusions

This study shows that an analysis of this type of system can be done by this method. The limitations are few and do not appear to seriously handicap the study. Limitations do occur in the cases where a generator needs to be represented by an induction motor or an induction generator. These conditions are extreme, but if their effects are deemed important, a proper analysis of the necessary data on the machine should make a representation of these conditions possible.

The time to make the studies would be another factor in evaluating this method. The original setup time on the computers was 2 days. The time for a run averaged about 90 minutes. However, both of these times could be shortened under the proper conditions. A considerable part of the setup time was used in calibrating the various circuits on the electronic differential analyzer. If the various circuit elements could be retained from one setup to the next, this calibration would not be necessary. Only a quick check of it would suffice.

The use of a wattmeter to check the power intake on the motoring machine would speed up all of the runs of this type, since a visual indication of this point would save much time consumed in adjusting the generators. A method of giving a direct reading of the excitation voltage would also speed up the time for various runs.

A complete evaluation of this method would involve a direct comparison with other methods of such factors as time and the cost of the necessary equipment.

Appendix. Representation of Voltage and Speed Regulator

Voltage Regulator

To represent the regulator with the electronic differential analyzer, the function of the regulator must be spelled out analytically. In this case, the circuit of the voltage regulator indicates that the analytical expressions which define the correction voltages supplied by the reactive-load equalizing circuit are as follows:

1. Correction voltage for the unfaulted generator

$$E_{u0} = -j\frac{M}{n}\left[I_{fta} - \frac{I_{ua}}{(n-1)}\right]$$

2. Correction voltage for the faulted generator.

$$E_{f0} = \frac{j(n-1)M}{n}\left[I_{fta} - \frac{I_{ua}}{(n-1)}\right]$$

$$E_{f0} = (1-n)E_{u0}$$

The electronic-differential-analyzer circuits to represent the foregoing equations are shown in Figs. 4(A) and (C). The value of E_{u0} is developed first and then the value of E_{f0} as $(1-n)E_{u0}$. In Fig. 4(A), the three sequence components of the phase A current signals are added. It should be noted that by using the opposite polarity on the transformer supplying these quantities (see Fig. 2), the output of amplifier *1* represents the positive value of I_{fta}, and the output of amplifier *2* represents the negative value of I_{ua}. Fig. 4(C) shows how these outputs are fed to the potentiometers, which represent the constants in the equations, and then summed as the input to amplifier *7*. The input to amplifier *7* is thus $[M/n][I_{fta} - I_{ua}/(n-1)]$. Amplifiers *7* and *8* each rotate this quantity -45 degrees for a total rotation of -90 degrees. This represents the $(-j)$ in the equation. The output of amplifier *8* represents the correction voltage E_{u0}. Amplifier *9* multiplies the value of E_{u0} by the constant $(1-n)$ to give the value for E_{f0}.

To make use of these correction voltages E_{u0} and E_{f0}, it is necessary to have the phase voltages of the load bus. Fig. 4(B) indicates the method of rotating the positive- and negative-sequence voltages so that they may be combined to produce phase voltages. Amplifiers *12*, *13*, and *14* in Fig. 4(D) are summing amplifiers which combine the three sequence quantities to give the three phase quantities. Amplifiers *10* and *11*, in this same figure, combine the phase A voltage with each of the correction signals. Fig. 6 is a phasor diagram showing the relationship of the various currents and voltages used in the voltage-regulator circuit. This diagram represents a typical case where $n=2$ and an unbalance in reactive power exists. The dashed lines on this diagram represent the distorted set of voltages that determine the regulator sensing signal during the reactive unbalanced condition. Fig. 4(E) shows how this same set of voltages is applied to the bridge rectifiers in the same manner as in the actual regulator. The outputs of these rectifiers are the regulator signals to the exciter, E_{fr} and E_{ur}.

The voltage regulator in this particular case has two limitations, maximum and minimum excitation voltage. This excitation voltage can be described by

$$e_d = e_q + I_1(X_d - X_q)\sin\theta$$

The phasor diagram representing this equation is shown in Fig. 3. Angle θ is the angle between the voltage e_q and the current I_1. Since the voltage behind quadrature

reactance e_q is used as the board source, the excitation voltage e_d can be determined by evaluating the quantity, $I_1(X_d - X_q)$ sin θ. As an aid in evaluating this quantity, the circuit shown in Fig. 4(D) involving sign changers *15* and *16* was used. The amplification in these two stages serves to put this quantity on the same voltage base as the voltage e_q. The maximum and minimum excitation voltage used in this case was 4.85 per unit and zero.

Speed Regulator

The circuit of the speed regulator indicates that the formulas for the power indication and the regulator output signal are:

1. Power indication for the faulted machine

$$P_{fr} = |0.324 I_{fta} + V_A| - |0.324 I_{fta} - V_A|$$

2. Power indication for the unfaulted machine

$$P_{ur} = \left|\frac{0.324 I_{ua}}{(n-1)} + V_A\right| - \left|\frac{0.324 I_{ua}}{(n-1)} - V_A\right|$$

3. Output signal to the faulted machine

$$S_{fr} = \frac{(n-1)}{n}(P_{fr} - P_{ur})$$

4. Output signal to the unfaulted machine

$$S_{ur} = -\frac{1}{n}(P_{fr} - P_{ur})$$

The electronic-differential-analyzer circuit to represent these equations is shown in Fig. 4(F). Sign changer *17* is used only to change the sign of the bus voltage. Sign changers *18* and *19* are used to give the power indication P_{fr} of the faulted machine. Sign changers *20* and *21* are used to give the power indication P_{ur} of the unfaulted machine. The input potentiometers and resistors of these latter four sign changers are used to set the coefficient of the current and to put it on the same voltage base as the bus voltage. The diodes in the feedback circuit introduce the absolute magnitudes. Amplifier *22* sums up these power indications, amplifies them to represent the constant $(-1/n)$, and gives an output which is the power-regulating signal S_{ur}. Amplifier *23* multiplies S_{ur} by $(1-n)$ to give an output which is the power-regulating signal S_{fr}.

References

1. Negative-Sequence Voltage Detection on Aircraft A-C Electric Systems, R. D. Jessee, P. F. Boggess. *AIEE Transactions*, vol. 75, pt. II, July 1956, pp. 171–75.

2. Electrical Transmission and Distribution Reference Book (book), Westinghouse Electric Corporation, East Pittsburgh, Pa., fourth edition, 1950, p. 175.

3. Power Equalizer Systems for Aircraft Alternators, John A. Granath, Albert K. Hawkes. *AIEE Transactions*, vol. 72, pt. II, Sept. 1953, pp. 209–17.

4. The Effects of Abnormal Conditions on Aircraft, Parallel A-C Power Systems, S. C. Caldwell, A. J. Wood. *Ibid.* (Jan. 1954 section), pp. 379–86.

5. Analog Computer Methods Applied to Steady-State A-C Problems, H. A. Kahle, H. M. McConnell. *Ibid.*, vol. 75, pt. II, Nov. 1956, pp. 279–82.

6. Symmetrical Components (book), C. F. Wagner, R. D. Evans. McGraw-Hill Book Company, Inc., New York, N. Y., 1933.

LATE DISCUSSION

The following discussion was received too late to be included with the paper.

Protecting A-C Motors with Low-Voltage Air Circuit-Breaker Series Trip

Discussion of paper 56-652 by F. P. Brightman, R. R. McGee, and P. J. Reifschneider, published in Applications and Industry, July, 1957, pages 114–19.

E. P. Ries (Shell Chemical Corporation, Martinez, Calif.): My comments are best presented by first briefly summarizing the electrical systems in our newer petro-chemical plants and then telling how the information presented in the paper affects us. Plants considered are characterized by continuous process operations. Most equipment drivers range from 3 through 75 hp (horsepower) with a few at 100 hp and larger. Continuity of service to most is critical.

Single- or double-ended selectively co-ordinated unit substations supply all 440-volt loads. Motors of 3 through 75 hp are controlled by conventional general-purpose control centers supplied through a feeder circuit breaker. Motors of 100 through 150 hp are controlled by individual large air circuit breakers.

This is in contrast to the older systems where unit substations were equipped with manually operated instantaneously tripped feeder circuit breakers. Motors of 3 through 75 hp were controlled by explosion-proof switchracks, each protected by a molded-case main circuit breaker. Motors of 100 hp and larger were controlled either by means of high-voltage contactors or power circuit breakers.

This means that we now expect the same selective tripping or co-ordination with the low-voltage system that we previously expected only from our higher voltage system.

Associated with this concept are the following problems which we feel deserve added attention:

1. Motor heating curves to permit proper application of trips are seldom available or obtainable.

2. Width of the trip tolerance band in some cases makes proper application either difficult or impossible.

3. It is necessary to obtain special factory calibration when close co-ordination is required.

4. Test equipment for an initial field calibration check is not available at an economical price.

5. For the same reason it is not possible to check the calibration at routine intervals as is done with a relay.

6. It would not always be possible to reset the trip properly in the field even if the necessary test equipment were available.

I should like to suggest that effort be directed toward the following:

1. Narrower trip tolerance bands.

2. Field test exuipment that can be justified by industrial users or redesign of trip units for use with doughnut-type current transformers so that normally available equipment can be used.

3. Provisions for complete field adjustment of trip units.

4. Publication of heating curves by motor manufacturers.

Power Apparatus and Systems—October 1957

Conference Papers Open for Discussion

Conference papers listed below have been accepted for AIEE Transactions and are now open for written discussion until December 26. Duplicate double-spaced typewritten copies of each discussion should be sent to Edward C. Day, Assistant Secretary for Technical Papers, American Institute of Electrical Engineers, 33 West 39th Street, New York 18, N. Y., on or before December 26.

Preprints may be purchsed at 40¢ each to members; 80¢ each to nonmembers if accompanied by remittance or coupons. Please order by number and send remittance to:

AIEE Order Department
33 West 39th Street
New York 18, N. Y.

57-777 Dynamic Representation of a Hydraulic Constant-Speed Drive .....Smith, Jr.
57-778 Optimizing Control Systems.....Cosgriff, Emerling
57-1037 Analysis of Multiple Sampler Systems with Finite Pulse Width (Open Loop).....Farmanfarma

AIEE PUBLICATIONS

	Member Prices	Nonmember Prices
Electrical Engineering		
Official monthly publication containing articles of broad interest, technical papers, and three news sections: Institute Activities, Current Interest, and Electrical Engineering Education. Automatically sent to all members and enrolled students in consideration of payment of dues.		annually $12* per year
		Single copies $1.50
* Subscription price and $1.00 extra for foreign postage both payable in advance in New York exchange.		
Bimonthly Publications		
Containing all officially approved technical papers collated with discussion (if any) in three broad fields of subject matter as follows:	annually	annually
Communication and Electronics	$2.50†	$5.00‡
Applications and Industry	$2.50†	$5.00‡
Power Apparatus and Systems	$2.50†	$5.00‡
† Members may elect to receive any one of the three bimonthly publications in consideration of payment of dues without additional charge.		
‡ Subscription price and 50 cents extra for foreign postage both payable in advance in New York exchange.		
Single copies may be obtained when available.	$1.00 each	$1.00 each
AIEE Transactions		
An annual volume in three parts containing all officially approved technical papers with discussions corresponding to six issues of the bimonthly publication of the same name bound in cloth with a stiff cover.		
	annually	annually
Part I Communication and Electronics	$3.00	$ 6.00**
Part II Applications and Industry	$3.00	$ 6.00**
Part III Power Apparatus and Systems	$3.00	$ 6.00**
Annual combination subscription to all three parts (beginning with vol. 74 for 1955).	$9.00	$12.00***
Annual combination subscription to any two parts.		$10.00***
** Subscription price and 75 cents extra for foreign postage both payable in advance in New York exchange.		
*** Subscription price and $1.00 extra for foreign postage both payable in advance in New York exchange.		
Electrical Engineering and Transactions		
An annual combination subscription to both publications (effective April 1, 1955).		$20.00§
§ Subscription price and $2.00 extra for foreign postage both payable in advance in New York exchange.		
AIEE Standards		
Listing of Standards, test codes, and reports with prices furnished on request.		
Special Publications		
Committee reports on special subjects, bibliographies, surveys, and papers and discussions of some specialized technical conferences, as announced in ELECTRICAL ENGINEERING.		

Discount 25% of above nonmember prices to college and public libraries. Publishers and subscription agencies 15% of above nonmember prices. For available discounts on Standards and special publications, obtain price lists from Order Department at Headquarters. Send all orders to:

Order Department
American Institute of Electrical Engineers
33 West 39th Street, New York 18, N. Y.